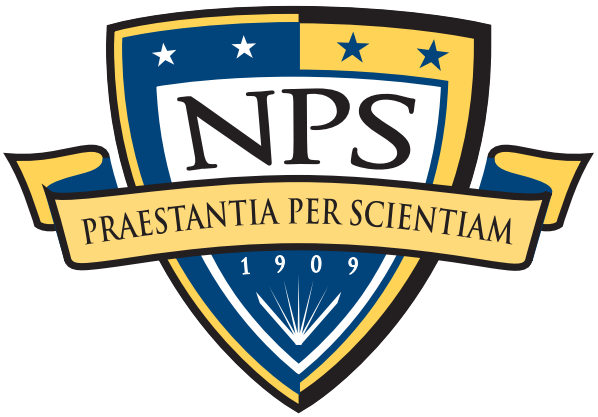

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AN ANALYSIS OF THE APPLICABILITY OF FEDERAL LAW REGARDING HASH-BASED SEARCHES OF DIGITAL MEDIA

by

Matthew B. Roy

June 2014

Thesis Advisor:	Simson Garfinkel
Second Reader:	Dorothy Denning

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	06-20-2014	Master's Thesis 01-15-2014 to 06-20-2014

4. TITLE AND SUBTITLE

AN ANALYSIS OF THE APPLICABILITY OF FEDERAL LAW REGARDING HASH-BASED SEARCHES OF DIGITAL MEDIA

5. FUNDING NUMBERS

6. AUTHOR(S)

Matthew B. Roy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views expressed in this document are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol Number: N/A.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

The Fourth Amendment of the United States (U.S.) Constitution limits the ability of the government to search U.S. persons without cause or justification. The application of the Fourth Amendment to digital forensics search techniques is still evolving. This thesis summarizes current federal law and recent judicial rulings that can apply Fourth Amendment doctrine to current digital forensics techniques. It uses three hypothetical scenarios to show how current law could be applied to new techniques now under development: the use of sector hashes to find traces of digital contraband; the use of random sampling to rapidly triage large digital media; and the use of similarity functions to find documents that are similar but not identical to target documents.

14. SUBJECT TERM

Digital forensics, hash-based search, sector hashing, random sampling, similarity matching, Fourth Amendment, federal law, search and seizure, warrant search, consent search, border search.

15. NUMBER OF PAGES 107

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

AN ANALYSIS OF THE APPLICABILITY OF FEDERAL LAW REGARDING HASH-BASED SEARCHES OF DIGITAL MEDIA

Matthew B. Roy
Lieutenant Commander, United States Navy
B.S., United States Naval Academy, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
June 2014

Author: Matthew B. Roy

Approved by: Simson Garfinkel
Thesis Advisor

Dorothy Denning
Second Reader

Peter Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Fourth Amendment of the United States (U.S.) Constitution limits the ability of the government to search U.S. persons without cause or justification. The application of the Fourth Amendment to digital forensics search techniques is still evolving. This thesis summarizes current federal law and recent judicial rulings that can apply Fourth Amendment doctrine to current digital forensics techniques. It uses three hypothetical scenarios to show how current law could be applied to new techniques now under development: the use of sector hashes to find traces of digital contraband; the use of random sampling to rapidly triage large digital media; and the use of similarity functions to find documents that are similar but not identical to target documents.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

CBP Customs and Border Protection

CD Compact disc

CDT Comprehesive Drug Testing, Inc

DC District of Columbia

DEA Drug Enforcement Agency

DMV Department of Motor Vehicles

DNA Deoxyribonucleic acid (the molecular basis of inherited genetics)

DOD Department of Defense

EFF Electronic Frontier Foundation

FBI Federal Bureau of Investigation

FIPS Federal Information Processing Standard

FTK Forensic Tool Kit

ICE Immigration and Customs Enforcement

ID Identification

IP Internet Protocol

ISP Internet Service Provider

KFF Known File Filter

MD5 Message Digest algorithm 5 (an cryptographic hash algorithm used to "fingerprint" digital documents)

MB megabyte

MiB mebibyte (1,048,576 bytes)

NCMEC National Center for Missing and Exploited Children

NIST National Institute of Standards and Technology

NPS Naval Postgraduate School

NSRL RDS National Software Reference Library Reference Data Set

NTFS New technology file system

P2P Peer-to-peer

PIN Personal Identification Number

RFC Request For Comment (an Internet standards document)

RAM Random Access Memory

SD Secure Digital

SHA-1 Secure Hash Algorithm-1 (another cryptographic hash algorithm)

SGCCCU Spanish Guardia Civil Computer Crime Unit

TB Terabyte

U.S. United States

USC United States Code

USG United States Government

Acknowledgements

I would like to thank my thesis advisor, Dr. Simson Garfinkel, for his outstanding guidance and mentorship. You always made me a priority and found time for me, despite the distance, whenever I needed it. I would not have completed my thesis without your support, expertise, and encouragement. I will always be grateful.

I would also like to thank Dr. Dorothy Denning for being my second reader and an invaluable resource here in Monterey. Thank you for your insight and constructive feedback.

There are several people from NPS that I would like to thank. I would like to thank VADM (ret) Daniel Oliver for the opportunity to come to NPS. It truly was an honor to serve as your aide. Thank you for your leadership, advice, and the sea stories. I would also like to thank VADM Jan Tighe for giving me the opportunity to complete my master's education at NPS. I would also like to say thank you to the staff that was always supportive when I worked up on the Mezzanine and during my time as a student.

Finally, thank you to my wife, Susan, and my three boys, Nathaniel, Brayden, and Zachary. Susan, thank you for all your support, strength, and encouragement over the course of my career to date and in the years to come. To my boys, thank you for being my inspiration and motivation.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

As technology continues to advance, becomes cheaper, and plays a more prominent part in the everyday lives of people, the role that digital forensics plays in criminal investigation will continue to grow. Where a criminal may have once kept physical copies of illegal media or a Rolodex with contacts, all of those things are now stored in digital format on computers, smart phones, tablets, or on servers in the cloud.

Current digital forensics techniques often use file hashing to search for illegal content. This technique reads a file from a physical disk drive or a disk image, computes the file's hash value using a cryptographic hash algorithm, and then searches for the hash value in a database containing the hash values of known contraband files. While this technique has been very successful to date, it can miss data matches because any modification to the suspect's file will result in a changed hash value–a value that will probably not be in the hash value database.

This thesis explores some of the legal implications of three new forensics techniques: sector-based hashing, random sampling, and similarity matching. These techniques can match data fragments or slightly modified data found on digital media with a database of known values allowing searches for partial matches and similar information. They also allow for faster searches based on random sampling of the subject media. These techniques thus allow for processing larger datasets and finding data that would have otherwise been missed by investigators.

Sector-based hashing focuses on the blocks of data that a file is divided into when saved to digital media. Instead of hashing the whole file, same-sized pieces of the file are hashed to produce multiple digests for the same file. This makes it much easier to locate a file that may have been deleted and partially overwritten by the file system or modified by the user. Sector hashing has a higher false positive rate than traditional file hashing because the same data block can be present in many different files. Thus, sector hashes require more interpretation than traditional file hashes. However, since sector hashing is file sys-

tem agnostics, it can be parallelized, allowing faster searches that are more fine-grain than traditional file-based hash searches.

When time is a critical factor, sector hashing can be combined with random sampling to quickly determine with high probability if target content is on a device. Random sampling uses the same methodology as sector-based hashing but instead of analyzing all the data, randomly selected sectors are read, hashed, and the resulting value checked for in a database. Probability and statistics are used to determine how confident the search results are and can be adjusted depending on the amount of time available.

Similarity matching can be file or block based. The basic premise is to find data that are similar to other types of data using a similarity function, where the meaning of the word "similar" depends on the specific function that is employed. Similarity matching techniques might allow for a broad range of information to be found that is currently being missed by the exact matching that file hashing requires. Although tools that employ similarity matching are widely available, these tools have not been widely adopted by digital forensics practitioners, perhaps a result of their high false positive rate when compared with traditional hash-based searching.

The legal implications of sector hashing, random sampling, and similarity based searching are the topic of this thesis. We know of no reported case in the U.S. legal system in which investigators made use of any of these techniques. U.S. courts generally do not issue abstract rulings on techniques and procedures unless they are used in actual cases that are being decided. Thus, the law always lags the technology. This thesis therefore attempts to apply current legal standards to these new techniques.

These techniques promise better searches in less time while keeping a person's right to privacy intact. However, their application in the future might be limited by current law and precedent.

The search methods described in this thesis have use beyond federal law enforcement. For example, these techniques can be applied to Department of Defense (DOD) operations including internal investigations, information assurance, counter intelligence, and system monitoring. DOD use may bring larger legal questions into play, such as the laws of other

countries, status of forces agreements, treaties, and customary international law. These questions are beyond the scope of this thesis.

This thesis examines three types of searches that can be executed by law enforcement officials: warrant searches, cross-border searches, and consent searches. Each has a very specific definition from a legal perspective. A hypothetical scenario will be presented and analyzed to explore how the current law would be applied in each of these cases involving the new forensics techniques described above.

- The first scenario involves a suspect giving consent for his vehicle to be searched because it matched a description of a vehicle connected to a kidnapping and murder. The investigators find a cell phone in the car that contains a micro Secure Digital (SD) card. A sector hash image is created and sent to a crime lab for analysis where sector hashes are identified that connect the SD card to the crime.
- The second scenario involves a border crossing search by Customs and Border Protection (CBP). In this scenario a United States citizen has the contents of digital media he is transporting across the U.S.–Mexico border examined for illegal content using random sampling.
- The third scenario involves an investigation into a member of a criminal organization that is engaged in credit card fraud. Federal law enforcement execute a search warrant on a suspect based off information from a previous arrest. The warrant allows for search of the suspect's computers and external drives which hold large amounts of data. Similarity matching is used to attempt to find documents similar to the ones from previous searches involving the same ring.

To make the most effective use of these scenarios, this thesis presents background material on digital forensics technology and U.S. law. Chapter 2 describes cryptographic hash algorithms and how they are currently used in digital forensics. The scenarios go beyond the techniques currently employed by forensic analysts, so this chapter also examines current research, including sector-based hash searches, random sampling, and similarity searches. Chapter 3 provides definitions of the three types of searches conducted by law enforcement officials that are examined in this thesis. It also presents pertinent federal law regarding searches. Chapter 4 summarizes what has been researched about the applicability of federal law to digital forensics to date. Chapter 5 presents the three hypothetical scenarios,

showing how current laws, statues, and precedence would apply to the new forensics techniques examined in Chapter 2. Chapter 6 presents questions that current legal approaches don't answer, discusses how those questions could be answered, and explores the issues that arise especially in regards to privacy along with the conclusion and recommended future work.

CHAPTER 2:
Technical Background

One of the primary tasks for the forensic analyst is making sense of data extracted from digital media. The data on digital media tell a story and will hopefully allow for the reconstruction of past events, allowing the analyst to answer the questions of who, what, where, when, why, and how. The goal is to find evidence or clues relevant to an investigation and be able to present that evidence in a court of law.

This chapter provides technical background on cryptographic hash algorithms and similarity functions. Cryptographic hash algorithms are a primary tool used by forensic analysts and the use of hashing is at the core of the first and second scenarios presented in this thesis. This chapter also discusses the use of similarity functions in digital forensics, which are the basis of the third scenario.

2.1 Cryptographic Hash Algorithms

A hash function is any algorithm or method that accepts, as input, data of any arbitrary size and produces a fixed length output called a *hash value* or *digest*. The output of a hash function is deterministic. That is, the hash value does not change if the same input is provided to the same hash function. These methods should allow for ease of computation and the result of the algorithm be something distinct that only the input could have produced.

Cryptographic hash algorithms are a specific type of hash algorithm that incorporate several key characteristics that make them suitable for use in applications involving computer security and forensics: the hash value should reveal nothing about the original message and any change to the original message should result in an unpredictably different hash value. In addition, cryptographic hash algorithms should exhibit Kerckhoff's principal in that the description of the algorithm should be publicly known and not require any secrecy to be secure [1].

2.1.1 Common Cryptographic Hash Algorithms

Message Digest algorithm 5 (MD5) and Secure Hash Algorithm-1 (SHA-1) are the two most common cryptographic hash algorithms used in digital forensics today. They are popular because their algorithms are well understood, their hash values are large enough to allow for a low probability of a collision, and they are relatively fast to compute.

MD5 was developed by Rivest and standardized in Request For Comment (RFC) 1321 [2] for use in digital signature applications. The algorithm produces a 128-bit hash value for inputs between 0 and $2^{64} - 1$ bytes in size. The algorithm works by breaking its input up into 512-bit blocks which are used as input into four rounds of calculations. The output at the end of these calculations is added to a set of four 32-bit states that when concatenated result in the function's value.

SHA-1 was the first member of the Secure Hash Algorithm family. They were published by the National Institute of Standards and Technology (NIST) and their formal specification can be found in the Federal Information Processing Standard (FIPS) Publication 180 [3]. These hash algorithms produce values ranging from 160 to 512 bits depending on the variant used. They were designed to be used with digital signature algorithms and keyed-hash message authentication codes. These algorithms work in a manner similar to MD5. Where these hash functions differ is in the size of the input blocks, the number and nature of the computations completed, and the size of the resulting hash value.

2.1.2 Security of Hash Functions

To analyze the security of cryptographic hash functions, cryptographers have identified three specific proprieties that the functions need to exhibit [1]:

- **Pre-image resistance:** Given the hash value of a message, it is computationally hard to find the message that produced the hash value. This is also referred to as the one-way property.
- **Second pre-image resistance:** Given a specific message, it is hard to find another message that produces the same hash value as the first. This is also referred to as weak collision resistance.
- **Collision resistance:** It should be computationally hard to find any two messages that produce the same hash values.

The *Random Oracle Model* [4] is a way to mathematically model how well a strong hash algorithm works. Under this model, a function called an oracle produces a random value every time it is presented with a new input. When it is presented with a previously seen input it provides the result that it provided on that previous time.

Table 2.1 demonstrates two hash functions that appear to follow the random oracle model. In this example, the "i" in helicopter was changed to a "j" by changing just one bit using a hexadecimal editor. Changing this single input bit results in approximately half of the output bits being changed. But the example only demonstrates one of the three security properties. Pre-image resistance results from the fact that the hash values are smaller than the message. Because the hash functions result in an irreversible loss of information, there is no way to transform the hash value back into the original message. Second pre-image resistance is not demonstrated by the example; it may be possible to find another message that has a similar hash value. Collision resistance is also not demonstrated: it may be possible to find another input that has the same hash value.

Original Text:	The helicopter was cleared to land by the tower.
MD5:	e4f46c3a0850c6bace630085cdc0fe36
SHA-1:	7bcad51170fcffd5c81790bea7c9bff2f4b4199c
Modified Text:	The heljcopter was cleared to land by the tower.
MD5:	96a5e3d323d070096c6e763504432621
SHA-1:	67afa1e22fec1fd9957bfd16618d1419e7705528

Table 2.1: Example of Cryptographic Hash Functions

It is clear that there must be many collisions for any hash function. A collision occurs when two different inputs into a hash algorithm produce the same value. Since the value produced by a hash function is only so long (128 bits for MD5, 160 bits for SHA-1 for example), there is a huge but finite number of possible digests. At the same time there are an infinite number of possible different data streams that could be used as input. By virtue of the pigeonhole principal, two or more of the inputs must produce the same output [1].

There are two potential sources of collisions: chance occurrences and intentional events. If the output of a hash algorithm is randomly distributed, then the probability of a chance collision depends entirely on the size of the hash value and the number of documents hashed.

This probability can be computed straightforwardly using statistics used to solve the birthday paradox (below). Computing the chance of an intentional collision is more difficult. The probability can be calculating for known attacks against hash algorithms, but it (obviously) cannot be calculated for unknown attacks.

2.1.3 The Birthday Paradox

The birthday paradox is a way to examine the probability of collision resistance. It asks the general question: assuming that birthdays are randomly distributed, how many people need to be in a room to be confident with high probability that any two of them have the same birthday? In the worst case scenario, there would need to be 366 people in a room, 365 persons each with a different day of the year, and then one more to have a match or in other words a collision. The worst case is usually not what occurs though, in fact the chance of match increases for every new person added to the pool. Mathematically, the probability can be found using the following formula [1]:

If $P(n) =$ probability that n people have different birthdays.

$$P(n) = P(1) \times P(2) \times P(3) \times \cdots \times P(n-1)$$

$$P(n) = 1 \times (1 - \frac{1}{365}) \times (1 - \frac{2}{365}) \times (1 - \frac{3}{365}) \times \cdots \times (1 - \frac{(n-1)}{365})$$

$$P(n) = \frac{n! \times \binom{365}{n}}{365^n}$$

Probability two people share the same birthday would be $1 - P(n)$

Using the above equation, it can be shown that with just 57 people, there is a 99 percent chance that two people have the same birthday.

The same math can be used to find the probability of a hash collision assuming that the hash digests are randomly distributed. We can show that the odds are statistically very low, especially when dealing with hashes that are more than 100 bits. For example, MD5 produces a 128-bit digest. That means there are 2^{128} possible different digests:

$P =$ Probability of 50% chance of any collision of an MD5 hash digest

$$P = .5 = 1 - (n! \times \binom{2^{128}}{n} (2^{128})^n)$$

To have a 50 percent chance of finding any collision with a randomly distributed 128-bit hash function, a person would need to hash approximately $n = 1.8 \times 10^{19}$ different inputs [5].

2.1.4 Attacks Against MD5

Wang and Yu demonstrated vulnerabilities in the MD5 algorithm that allowed them to create multiple inputs that produce the same MD5 hash value [6]. The cost of the attack was significantly less than attempting to brute force a solution. The time to find the first 512-bit block of both messages was on the order of 2^{39} MD5 operations and the second block of both messages around 2^{32} MD5 operations.

This attack works by taking a data object such as a file and making small changes to it until it produces the same hash. This vulnerability makes it possible to have two documents that say different things but have the same MD5 value, allowing a digital signature to be moved from one document to another. As a result of the attack, MD5 has been deprecated for many computer security applications.

2.2 Hashing in Forensics

Digital forensics analysts have two primary uses for cryptographic hash algorithms: ensuring data integrity and searching files for known content.

2.2.1 Ensuring Device and File Integrity

Hashing is used often in digital forensics as a means to ensure the integrity of evidence. Ensuring data has not been tampered with plays an important role in establishing chain of custody. There can be no doubt that the evidence examined by the analyst and reported to the court is the same as the evidence seized at the scene of the crime. A digest produced

by a hash function of a data device, such as a hard drive image, can be taken when the data is first accepted as evidence and then always recomputed after any function or method is applied to the data to ensure the data has not been altered in any way.

Most forensic examinations begin with the examiner making a copy of the original media. This copy, called a *disk image*, is a sector-by-sector copy and contains the allocated files, file system metadata, and unallocated sectors. Proper procedure requires that the disk be accessed using a device called a *write blocker* that is designed to prevent any write operation initiated by the host computer from reaching the subject media. After the first copy is made, a second copy is made. The cryptographic hash values for both copies are computed and compared. If the two hash values match, then the disk images are believed to be accurate copies of the original media. This hash value is recorded to be used as part of the analyst's report. The original media is then stored and analysis proceeds using one of the copies. If any other party needs a copy of the media, they are given a copy of the disk image, which can then be re-hashed and have its hash value compared with the recorded value. This ensures that all parties analyzing the media are working with an authentic copy of the original extracted data.

This same technique can be applied at the file level as well, ensuring the integrity of files during analysis. During the forensics process, files are usually extracted from the disk image using forensics tools such as Sleuth Kit [7], Forensic Toolkit [8], or EnCase [9]. These tools feature methods to extract files using the file system or can carve files out of the image by looking for file header and trailer signatures by accessing the data directly. As with disk images, extracting the same files more than once, computing the hash value, and comparing the resulting digest allows analyst to assume the file is genuine and from the original media.

2.2.2 File-based Hash Searches

File-based hash searches compare the hash values of known files to the hash values of files that exist on digital media. These hash-based searches are typically used in two ways: to search for known content, or to exclude known content. In both cases, a database contains a listing of files with the index being the hash value of the file. The digital media, such as a disk image, has its files hash values calculated one at a time. The database is then

searched for a match to the hash value. When searching for known content, the analyst is then alerted to any matches. When excluding known content, an analyst may want to see all data except for content known to be irrelevant. In this case any matches are not presented to the analyst.

Hash-based searches make searching for a known file, such as an image of child pornography, significantly faster since it is much faster to search a database for a 160-bit hash value than to compare byte-for-byte millions of files to the file in question. There is one major disadvantage to this method of search: due to the way cryptographic hash algorithms work, if one bit in a file is different between two otherwise identical files, then the match will be missed because the hash values will be different. If a person is trying to hide a file from this type of search, all that person needs to do is add an extra byte to the end. The digest will be different and any search for the original file using hashing missed.

There are two major databases in use today that investigators often use to search for file-based hash value matches: the NIST National Software Reference Library Real Data Set (NSRL RDS) and the National Center for Missing and Exploited Children (NCMEC) suspected and known child pornography database. The NSRL is a collection of millions of traceable software files. The files come from various commercial and open sources and include metadata for each file. From these files a collection of digital signatures has been created to form the Reference Data Set (RDS), which can be used by law enforcement and other digital forensics groups to help determine which files may or may not be useful in the course of a search. To date there are no MD5 or SHA-1 collisions in the RDS according to NIST researchers [10].

The National Center for Missing and Exploited Children houses a database of known and suspected child pornography files [11]. The difference between a known and suspected file is that the victim has been identified for known files, meaning that the person's age at the time of the crime is known. In the suspected files, since the victim is unknown, the age at the time can only be inferred. NCMEC provides identification and analysis services to law enforcement officials using hash-based searches of their database.

2.3 New Uses For Hashing in Digital Forensics

Much of the research is hash-based forensics is focused on sub-file forensics and similarity matching. With sub-file forensics, instead of hashing whole files, a file is broken up into blocks smaller than the file itself and those blocks hashed. This can be combined with random sampling to expedite the search of digital media while maintaining acceptable levels of performance. Similarity matching attempts to address a weakness in all hash based forensics: that the hash values from the digital media must be an exact match to the hash values in the database. Similarity matching attempts to produce matches to known data by providing a probability that two data objects are similar.

2.3.1 Block Hashing

When hash functions are used in digital forensics, they are usually applied at the file level. When files are stored on digital media they are not allocated in a space that matches their file size. Files are broken up and stored block-by-block on the disk media. In older disk drives, this minimal block size was typically 512 bytes for a hard drive, but more and more media is transitioning to 4096 byte block sizes. Block hashing, also called sector hashing, can be used for both hash searches in a manner similar to file-based hash searches and for hash-based file carving.

Hashing blocks of a disk image instead of whole files has some distinct advantages. The hash algorithm is applied directly at the sector level. The data can be read directly off the media and hashed without access to the file system or without the need of file carvers attempting to piece together files based on signatures. Since the blocks are read directly, the drive doesn't need to jump around, reducing the number of seeks. It is also easy to parallelize since blocks do not need to be pieced together to form files. Using blocks also increases the probability of finding a match of deleted or damaged files as only one block of data needs to be found in order to be matched. However, block hashing has disadvantages as well. Block hash databases require significantly more space to store their hashes, as where a file only had one hash before, it may now have hundreds or even thousands depending on the file and block size. Also, block hash databases have the possibility of false matches, since two different files can contain identical blocks of data. (In particular, many files contain blocks filled with the Unicode NULL (U+0000) character.)

Applying sector hashing to hash-based searches in forensics is similar to file-based hash searches. As the sectors are hashed, the hash values are checked for a match in a database of hash values and the results are either alerted or suppressed depending on the type of search. This technique also lends itself to hash-based file carving. Traditional file carver programs look for signature header or trailer bytes of a particular file in the data stream of a disk image. As signatures are found, the carver attempts to carve out chucks of data and reassemble it. Carving allows for the searching of content that may have been deleted or hidden in the slack space of a digital media device or any other data that is not accessible via the file system. With a hash-based carver, as a sector is carved, its hash value can be searched for in a database. If a match is found, the sector is known to be from a particular file and a particular location in that file. As more sector hashes from that file are found, the file can be reassembled from the data.

Research by Foster et al. [12] examined the feasibility of using block-based hash searches. The authors hypothesized that comparing a block to a large data set and showing it was distinct would allow the forensic analysis to treat the block as if it was universally distinct for the purpose of proving a file did exist. They proposed that using a database of hash values generated from fixed sized blocks of data would provide a faster and more accurate means of analyzing any form of digital media for a set of target data. The authors hashed the data of three large corpora: Govdocs [13], OCMalware [14], and the NSRL RDS [15], and did an analysis to determine the number and nature of any hash value matches that the large collection of data produced. They also tested several database implementations to determine if currently available databases could support the number of look-ups needed to match the rate the data are being read. Their research showed that the majority of blocks were distinct and that matches that were common were due to the same block existing in many files (such as the block of all zeros) or in some cases, the reuse of code in malware. Their research also showed that there was not a significant loss of precision from using a 4096-byte block size verses a 512-byte block size.

2.3.2 Random Sampling

The size of data storage available today is large even on consumer laptops and personal computers. Random sampling is a technique to help combat the ever increasing size of digital storage. It combines block-based sector hashing with statistics and probability the-

ory. Target data is broken into blocks that match the sector size of the digital media, those blocks have their hash value calculated and the hash values put into a database. Randomly selected sectors are then read off the digital media, hash values calculated, and their hash values checked against the database for a match. The sectors are read directly so it is file system agnostic. A match indicates that the target file does or did exist on the media at some point. This can be done with high confidence that the target data will not be missed while costing significantly less time than reading and hashing every sector on the digital media.

The technique takes advantage of the mathematical properties of sampling without replacement, also known as the Urn problem [16]. The Urn problem presents a scenario where an urn is filled with two types of balls, red and black. It is used to demonstrate calculation of the probability of finding a ball of a certain color giving that so many are taken out of the urn. Since the balls are randomly distributed and not replaced, the probability of removing a certain color ball can be shown to be high compared to the total number of balls in the urn. A variation of the equation for calculating that probability in the Urn problem is used to calculate the probability of missing a block of target data using random sampling [17]:

$$P = \prod_{i=1}^{n} \frac{((N-(i-1))-M)}{(N-(i-1))}$$

where N is the number of sectors on the digital media

M is the size of the target data as a multiple of the sector size

and n is the number of sectors sampled

Research by Taguchi [18] explored a hash-based random sampling method that would balance a high probability of detection with speed. He wrote a program that would randomly read blocks of data off of a drive, calculate the hash value of the blocks, and check a database of hash values of target data blocks. He experimented with the size of the data block as a multiple of the sector size to determine which offered at least a 90 percent probability of detection while taking the least amount of time. The size of the block read at a single time was called the transaction size, and they developed a range of them to cover

digital media where the drive layout is known and when it is unknown. He showed that random sampling could provide a 90 percent confidence of finding one block of 10 mebibytes (MiBs) of target data in 26 minutes on a one terabyte (TB) hard drive.

2.4 Similarity Matching

When using any hash-based forensic search technique, the hash value of the data must match a hash value in the database being searched exactly. This is due to the nature of cryptographic hash algorithms and is desirable when searching for known content. There is increased interest in finding content that is not an exact match, but is similar to known content. At the same time, the amount of data to be searched continues to increase. A technology is needed that reduces the amount of data that a human needs to examine. Similarity matching attempts to address this area of forensics [19].

The concept of similarity matching is straight forward: two files, such as documents or pictures, are similar if a human being says they are similar. To the human being the context and meaning of the data is important. While this is intuitive for people, it is difficult to formalize so that it can be implemented with a computer. Similarity matching is built around similarity functions which use an algorithm to determine if two files are similar. This is typically done at the byte stream level so there is no interpretation of what the data really means.

Kornblum developed a technique that expanded on Rabin's data fingerprinting research [20]. He adapted a technique developed for spam filtering to digital forensics. His method incorporated context-triggered piecewise hashes, often referred to as *fuzzy hashes*, as the basis for the comparison of two data streams. Typically done at the file level, the data is read in and broken into pieces using a rolling hash function. Those peices are then hashed using a different hash algorithm. The results are concatenated and base-64 encoded to produce a signature for the whole data object. Signatures of data objects are compared to each other by treating the whole fuzzy hash as a string and calculating the edit distance measure (the number of operations it would take to convert one fuzzy hash to the other). The technique assumes that related files would not require as many editing operations and therefore produce a higher confidence that they are similar. The key component of this method is determining where to break the file up into pieces. Some previous implementations used a

fixed length piece size. In this method the rolling hash reads in several bytes at a time and calculates a pseudo-random value for that piece. When the value equals the remainder of a constant then a trigger has been met and the previous bytes treated as a piece. The process then starts over until the whole object is processed. The constant is chosen based on the length of the data because the final hash is designed to be 80 characters in length. The tool developed incorporating this method is SSDeep [21].

Roussev's research took a different approach termed similarity digests [19]. His method attempts to locate 64-byte chucks of data, which he termed features, that have the lowest empirical probability of being encountered by chance. This is accomplished by calculating a normalized Shannon entropy measure that places any feature into one of 1000 classes of equivalence. These features are then hashed and placed into a Bloom filter. As a filter reaches capacity a new one is created until all the hash values from the designated features are added. The similarity digest is the sequence of all of these Bloom filters. Comparison is found by measuring the Hamming distance between the sets of Bloom filters of two data objects and averaging the number of matches together. The program developed incorporating this technique is SDHash [22].

Shields has developed a method designed to find similar content in text files [23]. The tool, called SDText, was first published in October 2012. The method takes a set of base files, tokenizes them based on several user selectable options (such as individual words, individual lines, words in a different language, or combinations of different tokenizers) and creates a dictionary consisting of the tokens and their statistical significance. Digests are then calculated using the dictionary and the target files being examined for similarity using bit vector fingerprints. A digest for a file contains a bit array representing the presence of tokens in the file. Digests of the target files are compared using cosine similarity. Each bit in the array is treated as a vector in space. If the cosine of the angle of all the vectors falls within a specified range then the files are considered similar. The software is very flexible in that it offers the user the ability to tune all the parameters allowing individual experimentation and selection of the features which produce the best matches depending on the needs of the user.

All of these methods are still relatively new. Both SSDeep and SDHash have been published in peer-reviewed literature [20], [24]. Empirical testing has been done to show that

they can produce varying levels of successful results [19], [25], [26]. To the author's knowledge, SDText has not been peer reviewed or had empirical testing results published. Their acceptance as a legitimate forensics technique depends on more scientific validation. By its very nature, similarity matching is much more inclined to producing higher false positive rates compared to traditional hashing techniques. This is to be expected as similarity can be subjective and these methods attempt to make the process objective and automated. Any data alerted as being similar would still need to be verified by an analyst. While these tools seek to reduce the analyst's workload through automation, questions remain as to the ability of these methods to find similar data and not miss important matches.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Federal Law

The law applicable to searches is rooted in the United States Constitution and is sourced from the United States Code and the precedence set from the rulings from court cases. The hypothetical scenarios specifically examine federal law, ignoring international and state law in order to maintain proper scope.

3.1 Types of Searches

The three types of searches examined are the kinds of searches that federal authorities typically execute. Each has specific requirements and limitations that are discussed below.

3.1.1 Warrant Searches

A warrant can be issued for the search of a person, property, item, or even information after probable cause has been established that evidence of a crime or illegal activity will be found when executing the warrant [27]. Probable cause is not absolute certainty. When dealing specifically with information on a computer there are several factors for the agents to consider:

- The warrant will typically need to specify that the search is for records or information meaning that the real interest is the data on the computer and not the computer itself (which might be the case if the computer were stolen property).
- The warrant will usually need to justify off-site examination of the data due to the amount of time it takes to forensically examine data on a hard drive.

Warrants do not need to specify how a search is to be conducted and therefore should not limit the techniques used in examining that data. A warrant is needed anytime a search will violate a reasonable expectation of privacy unless there is an exception to the warrant requirement such as when crossing a border or when consent to search has been granted.

3.1.2 Border Searches

A routine non-intrusive search is allowed without a warrant, probable cause, or reasonable suspicion at all border crossings regardless of whether the person or property is coming into or exiting the country and regardless of whether the person qualifies as a U.S. person [27]. Federal law enforcement authorities are authorized to search without a warrant due to their requirement of protecting the United States from contraband and other illegal property entering the country. This is not considered a violation of the Fourth Amendment of the Constitution because the searching of travelers crossing the border is considered reasonable [28]. In recent years, border crossings searches have been expanded to include the authority to search electronics such as laptops and portable hard drives.

3.1.3 Consent searches

Federal agents may search without a warrant or probable cause if a person consents to a search of a place or object voluntarily. Those executing the search must keep several factors in mind as the burden of proof that consent was given lies with the law enforcement officials. Those volunteering consent must be of proper age, intelligence, and physical and mental state [27]. Other factors need to be taken into account such as if the person was under arrest; and whether the person had been advised of his right to refuse consent. It is very important for investigators not to exceed the scope of the consent. Asking to search a room may not necessarily allow for search of the data on a computer in that room. The courts have not made a clear the line between when the scope of a search is exceeded especially in regards to data on computers when evidence of another crime is found. Agents must also remember that the person in question has the right to revoke his or her consent at any time. This can be especially problematic if digital media is removed for more detailed examination off-site.

3.2 Relevant Law

The Constitution of the United States and the United States Code are the basis for what is allowed during a search and seizure. This is true regardless of whether it is a search of one's physical possessions or the data stored on their computer. These laws seek to protect citizens from government officials being allowed to search whatever they want by providing a standard for what is required before a search can take place and for ensuring

the items or information searched is untampered with.

3.2.1 The Constitution of the United States

The Fourth Amendment of the Constitution, ratified as part of the Bill of Rights on December 15, 1791 states the following [29]:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

In the Colonial era, officials representing the King were allowed to execute general warrants allowing searches for evidence of any crime. The Fourth Amendment was written to specifically forbid those types of actions by the government. When a government agent enters a home a search occurs that violates the privacy of whomever lives there [30]. Justification must be provided as to why the search is being done that meets the standard of probable cause: a fair probability that contraband or evidence of a crime will be found in a particular place [27]. In order to prevent a general search of a person or property the amendment specifically states the warrant must describe the place to be searched and the person or things to be seized.

3.2.2 Rules of Criminal Procedure for Search and Seizure

The Federal Rules of Criminal Procedure are part of the United States Code Title 18 Appendix [31]. Rule 41 defines the regulations and restrictions governing federal agents when executing a search and seizure. A warrant can be requested by a federal law enforcement officer or an attorney representing the government. It must be issued by a magistrate judge or in some cases, a judge of a state court. Usually, that judge must have authority within the district for which the warrant is issued. The judge may issue a warrant after determining there is probable cause to search for or seize a person or property. Probably cause is determined after the judge has been presented with a signed affidavit, sworn testimony, or recorded testimony.

According to Rule 41 a warrant may be issued for any of the following [31]:

- Evidence of a crime.
- Contraband, fruits of crime, or other illegally possessed items.
- Property designed for use, intended for use, or used in committing a crime.
- A person to be arrested or a person who is unlawfully restrained.

The definition of property includes information and not just tangible items. In addition, when a warrant seeks electronic information the media itself may be seized or a copy of the media made. Unless specifically prohibited,, the media may be reviewed for pertinent information at a later time. Any time line placed on the warrant usually applies to the initial collection or copying of the data [31].

When the warrant is issued, it must be issued to a law enforcement officer. The warrant must specify the person or property to be searched and seized, and designate the judge to whom the warrant should be returned. Upon issuance, the warrant must be executed within 14 days and during the hours of 6AM to 10PM unless authorized otherwise. A full inventory of all items seized must be made. When electronic media is involved the officer is allowed to retain a copy of the information that was seized or copied [31].

3.2.3 Rules of Evidence

The Federal Rules of Evidence apply to all proceedings in all United States courts [32]. They are designed so that all court proceeding are fair, eliminate unjustifiable expense and delay, and promote the development of evidence law all toward the goal of ascertaining the truth and finding a just determination. They define the conditions and circumstances in which something may be admitted into a legal proceeding as evidence. The rules are very general as they attempt to cover all the different types of evidence that can be presented. The most pertinent rules from a forensics perspective, are the rules of Article IV, VII, and X.

Article IV defines relevant evidence and the limitations imposed on it. Evidence is relevant if [32]:

- It has any tendency to make a fact more or less probable than it would be without the evidence.

- The fact is of consequence in determining the action.

There are some exceptions. Relevant evidence is not allowed if it has an exception listed in the Constitution, a federal statue, the Rules of Evidence themselves, or any other rule the Supreme Court prescribes. Irrelevant evidence is not admissible under any circumstance. Evidence may also be excluded if it causes unfair prejudice, confuses the issue, misleads the jury or wastes time [32]. Any evidence found in the course of an investigation involving digital forensics must meet these requirements.

Article VII defines opinions and expert testimony. An expert witness is one who, due to their knowledge, experience, training, and education, can be considered an expert in their field. They are allowed to testify their opinion if [32]:

- The scientific, technical, or other specialized knowledge will help the person or persons deciding the case to understand the evidence or determine a fact.
- The testimony is based on sufficient facts or data.
- The testimony is produced by reliable principals and methods.
- These methods and principals have been reliably applied to the facts of the case.

Typically a forensic practitioner will be brought before the court to testify as to the findings of the analysis of digital media completed in the course of an investigation. This person will not only present facts about the data recovered but is allowed to state his opinion. Article VII has had significant changes due to *Daubert v Merrell Dow Pharmaceuticals, Inc* (p. 33) where the validity of expert testimony came into question.

Article X addresses the contents of writing, recordings, and photographs. While this was originally designed to address just these physical items it has been expanded to cover electronic versions of these documents. Specially, for electronically stored information, an original constitutes a printout or other output that can be read as long as it accurately reflects the information and that duplicates can be produced via electronic means as long as the technique used to make the copy accurately reproduces the original. Both are allowed and in some cases required in order to prove the content of the item [32].

3.2.4 Customs Duties

Title 19 Section 482 of the United States Code [33] is the law that authorizes customs officials to conduct searches. It authorizes the search of any person, property, vessel, or vehicle entering the United States and with reasonable cause, the seizure of any property that has been brought into the country illegally and detainment of the person who attempted to bring it in.

3.2.5 Wiretap Act

Title 18 Sections 2510 through 2522 of the United States Code [34] are the federal laws regarding the interception of oral, wire, and electronic communications. The laws in general prohibit the collection or interception of communications made via wires, cables, and electronic means unless authorization is specifically authorized by an appropriate judicial authority or the consent of one of the parties is obtained. This set of laws is commonly refered to as the Wiretap Act.

3.2.6 Pen Registers and Trap and Trace Devices Statute

Title 18 Sections 3121 through 3127 of the United States Code [35] are the laws regulating the use of pen registers and trap and trace devices. In general they are prohibited unless an exception applies. This exception usually applies to law enforcement and other government agents. These devices are used to record what types of communications people are conducting while keeping the content of the conversation secret such as phone numbers. The standard for obtaining authorization is lower than a wiretap which requires probable cause. When authorized, their use covers the decoding of any wired or electronic communication routing and addressing in order to determine the source of the communication.

3.2.7 Stored Wired and Electronic Communication Act

Title 18 Sections 2701 through 2712 of the United States Code [36] are the federal rules and regulations governing electronic communications where an electronic communication consists of "any transfer of signs, signals, writing, images, sounds, data, or intelligence of any nature transmitted in whole or in part by a wire, radio, electromagnetic, photoelectronic or photooptical system that affects interstate or foreign commerce" [34]. It defines the conditions by which service providers are required to cooperate with government agents such as

the voluntary or involuntary disclosure of records or communications and the requirements that the government must meet in order to gain proper access to those records.

3.3 Pertinent Cases and Rulings

The judiciary is responsible for the interpretation of constitutional and federal law. Through the trial, conviction, and the appeal process of real cases, courts set precedence about what is and is not acceptable as evidence which in turn establishes what is or is not allowed during a search. The following cases had significant impact on what is or is not allowed during a search especially when information on a computer is involved.

Table 3.1: Court cases mentioned in this thesis.

Year	Case	Covers	Page
1967	Katz v U.S. 389 U.S. 347	Expectation of privacy	26
1981	U.S. v Heldt 688 F.2d 1238	Limits on large scale warrant searches	27
1982	U.S. v Tamura 694 F.2d 591	Seizure of large quantities of data	28
1984	U.S. v Jacobsen 466 U.S. 109	Warrantless field testing	29
1987	Arizona v Hicks 480 U.S. 321	Probable cause and the plain view doctrine	30
1988	California v Greenwood 486 U.S. 35	Warrantless search of garbage allowed	31
1991	Florida v Jimeno 500 U.S. 248	Scope of consent searches	32
1993	Daubert v Merrell Dow Pharmaceuticals 509 U.S. 579	Standard of scientific evidence	33
1999	U.S. v Carey 172 F.3d 1268	Scope exceeded when evidence of another crime found	35
1999	U.S. v Upham 168 F.3d 532	Recovered deleted data within scope	36
2001	Kyllo v U.S. 533 U.S. 27	Expectation of privacy and new technology	37
2005	Illinois v Caballes 543 U.S. 405	Use of drug-sniffing dog allowed during traffic stop	37

Year	Case	Covers	Page
2005	U.S. v Brooks 427 F.3d 1246	Warrant should not limit how a search is conducted	38
2005	U.S. v Ickes 393 F.3d 501	Search of electronics allowed at border crossings	40
2006	U.S. v Grimmett 439 F.3d 1263	Computer searches may be extensive	40
2006	U.S. v Hill 459 F.3d 966	Broad seizure of data	41
2007	U.S. v Heckenkamp 482 F.3d 1142	Expectation of privacy and data	42
2008	U.S. v Arnold 533 F.3d 1003	Border search of computers	43
2008	U.S. v Cartier 543 F.3d 442	Hash value matches establish probable cause	43
2008	U.S. v Crist 627 F.Supp.2d 575	Warrantless hash search exceeded scope of private search	44
2008	U.S. v Giberson 527 F.3d 882	Computers are repositories	45
2009	U.S. v Comprehensive Drug Testing, Inc 579 F.3d 989	Guidelines for examination of data	47
2010	U.S. v Mann 592 F.3d 779	Hash search exceeded scope of warrant	48
2011	U.S. v Miknevich 638 F.3d 178	File name and hash value provide probable cause	49
2013	U.S. v Cotterman 709 F.3d 952	Forensics examination at border requires reasonable suspicion	50

3.3.1 Katz v United States 389 U.S. 347 (1967)

Definition of a reasonable and legitimate expectation of privacy

Charles Katz was convicted on an eight-count indictment that charged him with transmitting wagering information from a public pay phone in Los Angeles to Miami and Boston. Part of the evidence presented at his trial consisted of recordings of Katz's end of telephone conversations that were obtained by listening to an electronic listening and recording device attached to the outside of the phone booth. These recordings were obtained without

a warrant. Katz's appealed his guilty conviction stating the recordings were obtained in violation of the Fourth Amendment, but the Ninth Circuit Court of Appeals affirmed the conviction stating that there was no physical entrance into the area occupied by Katz. Believing the topic to be of constitutional significant, the Supreme Court granted certiorari and reviewed the case [37].

Up until this point surveillance had been viewed as not being a search and seizure unless physical penetration had occurred, such as planting a bug inside someone's home, because search and seizure only referred to physical and tangible property. Instead the Court decided that the protections of the Fourth Amendment extended to people and not just their property. With this in mind, the court determined that the recordings gathered by the federal agents did constitute a search and seizure which violated Katz's privacy. Having been presented with the actions the investigators took to limit the scope and duration of their surveillance to ensure their actions focused solely on Katz and the justification they used to survey him in the first place, the Court determined that any magistrate, having been briefed on these intended actions, would have issued a warrant authorizing the actions. The Supreme Court ruled that the conviction be overturned [37].

3.3.2 United States v Heldt 668 F.2d 1238 (DC Circuit 1981)

Searching among commingled records and the limitations on government agents when executing large scale search and seizures

Three search warrants were executed simultaneously for three premises owned by the Church of Scientology. One was in Washington, District of Columbia (DC) and the other two in Hollywood, California. The warrants were based on a 33-page sworn affidavit which detailed the results of a government investigation that alleged stealing of government documents, conspiracy to steal, and conspiracy to obstruct justice. The search involved over 200 federal agents due to the massive amount of records that needed to be searched. The warrant specified 162 descriptions of documents and at the end of the search and seizure over 23,000 documents had been recovered. Shortly after the warrant search concluded, the defendants filed a motion to suppress all the evidence gathered stating that the warrant search was actually a general search an violated the Fourth Amendment. The case was eventually appealed to the DC Court of Appeals [38].

The court recognized that strict bounds must be placed on a warrant in order to prevent a general search during the execution of a warrant. They determined there are three requirements when executing a search for potentially numerous documents [38]:

- **Adequate preparation:** The team conducting the search must be read into the terms that the warrant allows.
- **Obedience to area limitations:** It must be understood that the authority to search an area is limited by the specific places described in the warrant.
- **Restrictions on seizure of items not mentioned particularly in the warrant:** In general, only items particularly specified in the warrant may be searched and seized, however, evidence of other criminal activity may be seized as well when found as part of a valid search under the *plain view doctrine*. To be eligible under this doctrine, the agent must be lawfully in the location where the search is conducted, the item must be incriminating in nature to establish probable cause for its seizure, and it must be discovered inadvertently.

Using these guidelines, the DC Court of Appeals upheld the district court decision to not suppress all the seized documents. While some of the documents taken did fail to meet the plain view doctrine, none of those, as far as the court can tell, were entered into evidence. The motion to suppress the evidence was denied [38].

3.3.3 United States vs Tamura 694 F.2d 591 (9th Circuit 1982)

The seizure of large amounts of commingled data is allowed is rare cases

A seminal case involving an investigation into a large bribery, mail and wire fraud, conspiracy, and racketeering. The execution of the search warrant involved the seizure of a large, almost wholesale, amount of printed documents that exceeded the scope of the search because the information was commingled and would take too long to accomplish on-site. The court of appeals decided that in such rare cases, this is allowable but permission should either be sought ahead of time during application of the warrant, or the documents sealed and held until authority for a subsequent search can be obtained [39].

3.3.4 United States v Jacobsen 466 U.S. 109 (1984)

Defines limits of what is a search and seizure of property when conducting tests that can reveal contrband

A package, while in transit, was discovered ripped open by the employees of the Federal Express office at the Minneapolis-Saint Paul airport. In accordance with company insurance policy, the package was opened to examine the contents. Inside the cardboard box was a tube made of silver duct tape inside of which contained several baggies of a powdery white substance. A manager put the bags back in the tube, the tube back in the box, and called the Drug Enforcement Agency (DEA). An agent arrived on scene, examined the package and tubing, removed the plastic baggies, and conducted a field test with a trace amount of the substance.The powder tested positive for cocaine. Federal agents obtained a warrant to search the residence of the package destination. The residents were arrested and charged with possession of an illegal substance with intent to distribute. The defendants filed a motion to suppress evidence stating the warrant was based on an illegal search and seizure. The district court denied the motion, but on appeal, the 8th Circuit of the Court of Appeals reversed the decision stating that the validity of the warrant stood on the warrantless test of the white powder and that the testing was a significant expansion of the earlier private search by the employees. The Supreme Court realized that this ruling conflicted with another decision by a different court of appeals that had similar facts and granted certiorari to examine the circumstances [40].

The Supreme Court first stated that a parcel in transit is an "effect" as stated in the Fourth Amendment and the sender and recipient has a reasonable expectation of privacy. Even if a government agent is authorized to seize a package to prevent its destruction, that agent still requires a warrant to search the package. This status does not prevent law enforcement officials from acting if the parcel is examined by a private party and evidence of illegal activity discovered. The initial invasion by the private employees does not violate the Fourth Amendment regardless of how deliberate or unreasonable their search was because the Fourth Amendment does not apply to the actions of private actors. What determines whether the search exceeded the scope of the private search and was therefore unreasonable, were the actions taken by the field agent upon arrival and the facts known to the agent at the time. The Court decided that the initial inspection, to include removing the baggies

from the tube, did constitute a search and seizure but that those actions were reasonable. These actions resulted in the agent learning nothing more than had not already been learned through the testimony of the employees who had conducted the private search [40].

The second part of the decision centers around the field test. The Court needed to determine if this action constituted a search and if so did it infringe upon an expectation to privacy that society would consider reasonable. The Court referenced a decision by Congress to treat the interest in privately possessing cocaine as illegitimate. Based on this fact, using a field test which will only reveal a substance as cocaine is authorized because there is no legitimate privacy interest in illegal activities and the agent could not learn anything else about the substance that would compromise a legitimate interest [40].

The final part of the decision concerns the use of the substance in the test. Since a small amount of powder would be used to conduct the test, the respondent's do have a possessory interest since that amount would have been destroyed whereas before it was only temporarily deprived. The Court was required to assess the nature and quality of the intrusion, the advantage of the knowledge to be gained, and the person's interests. They found the actions reasonable based on the fact that it was virtually certain the substance was illegal and the trace amount of powder used to conduct the test. These actions were considered a valid warrantless seizure that was reasonable [40].

The decision was reversed and sets the precedent that techniques and methods which can only reveal an object of a search to be illegal or contraband are authorized with or without a warrant dependent upon the nature of the intrusion.

3.3.5 Arizona v Hicks 480 U.S. 321 (1987)

Police require probable cause when conducting search and seizure using plain view doctrine in a dwelling.

A bullet was fired through the apartment floor of Hicks injuring a person in the apartment below. The police arrived at his apartment in an attempt to search for the shooter, other victims, and the weapon used. Three weapons were found and seized. During the course of the search an officer noticed two nice sets of stereo equipment that seemed out of place compared to the conditions in the rest of the apartment. Suspecting that they might be

stolen, the officer recorded the serial numbers, moving the equipment in the process, and reported the numbers to his headquarters. Upon being informed the serial numbers matched equipment stolen during an armed robbery, the stereos were seized and Hicks indicted for robbery [41].

The trial court granted a motion to suppress evidence found as a result of recording the serial numbers and the Arizona Court of Appeals agreed. The courts stated that while the initial entry without a warrant was justified based on the circumstances surrounding the shooting, the act of obtaining the serial numbers was a separate search, unrelated to the original reason for entry, and thus violated the Fourth Amendment because it was done without a warrant. The state courts rejected the notion that the actions were justified under the plain view doctrine. After the Arizona Supreme Court denied review of the case, the State filed a petition to the Supreme Court. The Court granted certiorari [41].

The Supreme Court agreed that the actions of the officer constituted a separate search but rejected the opinion of the court of appeals that stated because the examination of the stereo equipment was for a reason different than the reason for entry that the search was unreasonable. The Supreme Court argued that this relationship always exists and can be justifiable under the plain view doctrine. The Court ultimately affirmed the decision of the Arizona Court of Appeals but for a different reason: under the plain view doctrine, probable cause is required when conducting a search and seizure of someone's dwelling or other places where a search without a warrant would normally be unreasonable. The Court pointed out that this is not always the case such as when the seizure is minimally intrusive and operating necessities require it as part of the mean to detect a crime [41].

3.3.6 California v Greenwood 486 U.S. 35 (1988)

A warrantless search and seizure of garbage left outside does not violate the Fourth Amendment

An investigator with the Laguna Beach Police Department received information that Billy Greenwood might be engaged in drug trafficking. She conducted surveillance of Greenwood's home noting that several vehicles make brief stops at the home late at night and in the early morning. She also followed a truck from the residence to another house that was suspected in narcotics trafficking. The investigator asked the neighborhood garbage

collector to pick up the bags outside of the Greenwood residence and turn them over to her. Searching through the trash, she discovered items associated with drug use. She applied for a warrant based on this information [42].

Upon execution of the warrant, police discovered cocaine and hashish. Greenwood was arrested on felony drug charges and posted bail. Approximately one month later, police inspected the garbage again after receiving reports of late night visits to the house. Again more evidence of narcotics use was found, a second warrant secured, and the follow on search led to discovery of more narcotics and evidence of trafficking. The California Superior Court dismissed the charges because the search of garbage without a warrant violated the Fourth Amendment and the California Constitution and probable cause to search the house would not have existed otherwise. The California Court of Appeals confirmed but pointed out that, under a California constitutional amendment, if the search was found reasonable under the Fourth Amendment, but still unreasonable under the California Constitution, that the evidence would be admissible. Since the court of appeals could assume that a garbage search was still unreasonable under both laws, they concluded that the decision stood. The State petitioned for the California Supreme Court but was denied [42].

The Supreme Court granted certiorari to specifically address whether the search of garbage without a warrant violated the Fourth Amendment. The Court decided that exposing garbage to the public defeats the claim to Fourth Amendment protection as it is well known that scavengers, animals, children, and other members of the public search through trash. In addition, the trash is placed outside so that a third party can collect it. As such, the Court concluded that society would not accept an expectation to privacy of trash left for collection as reasonable and reversed the decision by the California Court of Appeals [42].

3.3.7 Florida v Jimeno 500 U.S. 248 (1991)

When consent is granted for a vehicular search, the consent does extend to certain containers in the vehicle.

A Dade County police officer had heard Enio Jimeno arrange what appeared to be a drug transaction on a public telephone and decided to follow him. After making a right turn at a red stop light without stopping, he pulled Jimeno over. The officer informed Jimeno that he stopped him for a traffic violation but suspected him of carrying narcotics and asked for

permission to search the car. Jimeno consented to the search. Both Jimeno and his wife stepped out of the car. When the officer inspected the passenger side of the car he found a folded paper bag, opened it, and found a kilogram of cocaine. Both were charged with possession with intent to distribute cocaine [43].

The trial judge granted a motion to suppress the evidence found as a result of the consent search because his consent to a search of his car did not carry a specific consent to open the bag and examine the contents. The Florida District Court of Appeals and Florida Supreme Court agreed which established a rule that consent to a search does not extend to sealed containers within the general area that the defendant agreed to [43].

The case was granted certiorari by the Supreme Court which reversed the decision of the Florida Supreme Court. The key question they sought to answer was is it reasonable for the officer to extend the scope of the search to include closed containers in the car. The Florida courts believe that if a closed container is found while conducting a consent search that separate permission to search the container must be gained. They based their decision on a case where an officer had pried open a locked brief case in the trunk of a car while conducting a consent search. This was deemed unreasonable. The Supreme Court determined that the facts of this case were wholly different. A reasonable person would expect that narcotics are usually carried in some kind of container and since the officer had specifically told Jimeno he would be searching for narcotics, that this reasonableness would include containers which might bear drugs located in the car. The actions of the officer were reasonable, therefore the search was legal, and the evidence admissible [43].

3.3.8 Daubert v Merrell Dow Pharmaceuticals 509 U.S. 579 (1993)

The precedence for what is allowed to be admitted as scientific evidence is set.

The petitioners in this case, the parents of Jason Daubert and Eric Schuller, both born with birth defects, sued Merrel Down Pharmaceuticals. They believed a drug that the company produced named Bendectin, which was a prescription anti-nausea medicine, caused birth defeats in their children because the medicine was used during pregnancy. The suit was eventually moved to federal court on diversity grounds and after extensive discovery, Merrell Dow motioned for summary judgment stating that Bendectin did not cause birth defects in humans and that the petitioners would not be able to produce admissible evidence to the

contrary. As part of the motion, they submitted an affidavit of Doctor Steven Lamm, a well-credentialed physician and epidemiologist who had reviewed all literature available on the drug and birth defects encompassing 30 published studies involving 130,000 patients. No study had found Bendectin to be a cause of birth defects and the doctor concluded that maternal use of the medicine had not been shown to be a risk factor. The petitioners did not contest the published report but instead submitted testimony from eight different expects, each well-credentialed, that said a link did exist between use of the drug and birth defects. Their conclusions were based upon test tube and live animal studies [44].

The district court granted the summary judgment stating that scientific evidence is only admissible if the principle upon which it is based is sufficiently established to have general acceptance in the field to which it belongs and that the evidence presented by the petitioners did not meet that standard. The Ninth Circuit Court of Appeals agreed stating that the expert opinion must be based on techniques generally accepted as reliable by the scientific community and that the methods presented significantly diverged from procedures accepted by recognized experts in the field. The subject matter of the case had been brought up in other courts of appeal previously. The Supreme Court recognized that the general acceptance test was no longer sufficient for determining the admissibility of expert testimony and granted certiorari regarding the proper standard [44].

The Supreme Court examined two key items: *Frye v United States 293 F. 1013 (DC Circuit 1923)* [45] and the Rules of Evidence. *Frye v United States* saw the adoption of the general acceptance test that has stood as the primary method for determining the admissibility of evidence found with new scientific methods. The Rules of Evidence establish guidelines for who may provide expert testimony, the types of testimony they may provide, and the limitations on that testimony. The Court determined that the trial judge must be a gatekeeper in determining the admissibility of scientific evidence at the outset. The judge must base the decision on whether the witness will testify to scientific knowledge and whether the testimony will assist the judge or jury to understand the facts. To assist in the evaluation, the Supreme Court established several criteria to be considered [44]:

- The method or knowledge presented should be empirically testable.
- The technique or theory should have been subjected to peer review and publication.
- The potential or known rate of error for the method should be known.

- The methods should be generally accepted.

The Supreme Court determined that the district court and court of appeals had only applied the general acceptance test to the evidence concerning this case and remanded it for further review [44].

3.3.9 United States v Carey 172 F.3d 1268 (10th Circuit 1999)

While searching for evidence of one crime, law enforcement officials exceeded scope of search when evidence of another crime was found

The defendant, Patrick Carey, had originally been under investigation for possession and sale of cocaine. Police eventually obtain a warrant to arrest Carey after he made several purchases from undercover officers. During the arrest, officers found drug paraphernalia in plain view and obtained consent from the suspect to search the apartment. The search lead to discovery of additional drugs and two computers. Warrants were obtained, based on evidence found at the scene, to search for files on the computer which contained documentary evidence of the sale and distribution of controlled substances including names, telephone numbers, and receipts. During the course of the computer search, officers discovered an image containing what appeared to be child pornography. The officials continued to open subsequent image files under the suspicion that they contained child pornography as well. No warrant was obtained to expand the search to include these new files [46].

The defendant filed a motion to suppress the evidence of child pornography but the district court denied the motion. That decision was reversed by the Tenth Circuit Court of Appeals which stated that the scope of the warrant search had been exceeded and that the district court erred in not granting the motion in the first place. In this specific case, the testimony of the investigating detective played a key role. The court observed, based on the investigator's testimony, that after the first image of child pornography had been found that the detective did knowingly open the rest of the image files expecting to find more evidence of child pornography. The court decided that, at this point, the investigators should have applied for a second warrant to search for additional evidence of child pornography based on probable cause from opening the first image. The final ruling by the court stated that the plain view doctrine had been exceeded and that the evidence was found based on an unconstitutional general search. The court was quick to caution that while it felt this was the correct decision

because the investigators had switched the context of the search from drug related items to child pornography, this might not always be the case concering computer searches [46].

3.3.10 United States v Upham 168 F.3d 532 (1st Circuit 1999)

The recovery of deleted data does not exceed the scope of a search

As part of an undercover investigation, U.S. Customs agents were monitoring an internet chat room where a number of images depicting child pornography had been received. The agents contacted the Internet Service Provider (ISP) and traced the location of the computer which had sent the images. It belonged to Kathi Morrissey who lived in Costigan, Maine. Agents obtained a warrant and conducted a search of the home. A personal computer and a number of diskettes were seized as part of the search. Forensic analysts were able to extract over 1400 images from the digital media, some of which matched images viewed in the chat room. The majority of the images had been previously deleted. Further investigation lead the agents to determine that the homes inhabitants included Troy Upham, Morrissey's boyfriend. Evidence and Upham's own admission showed him to be the primary user of the computer and that he had been the person who sent the images. He was charged with four counts of interstate transport of computer graphical images depicting child pornography. Each count was tied to a separate internet transmission of the image. He was also charged with possession of child pornography. Upham filed a motion to suppress evidence derived from the search of the home. The district court denied the motion and he was tried and convicted. The denial of the motion to suppress evidence was appealed to the First Circuit Court of Appeals [47].

The appeal was based on two arguments: that the warrant was too generic and that the recovery of the deleted data was outside the scope of the warrant. On the first point, the court determined that the warrant as written was very specific and objective in its criteria. The court also addressed whether the warrant was too broad stating that the evidence which produced probable cause for the warrant justified a seizure and off-site search of the computer. In this case they also believed that the search of the computer and disks were no more obtrusive than the search of a home for a weapon or drugs. The court addressed the second point with a couple statements. The first was that the attempted destruction of something does not make it inadmissible, for example, a ransom note that is reconstructed

after it had been shredded and thrown away is still evidence. The court also pointed out that warrants do not require specifying how something is to be searched for [47].

3.3.11 Kyllo v United States 533 U.S. 27 (2001)

Use of new technology by law enforcement without a warrant can constitute an unreasonable search

An agent from the U.S. Department of the Interior suspected that marijuana was being grown in the home of Danny Lee Kyllo due to circumstances involving another investigation. Knowing that the indoor growth of marijuana required high intensity lamps, he used a technology new to law enforcement at the time, a thermal imaging camera, to covertly scan the house for high heat signatures. Based on the scans, tips from an informant, and utility bills that were well above average for a house of its size, the agent requested and was issued a warrant to search the home for drugs. Upon execution of the warrant, more than 100 marijuana plants were found and Kyllo was arrested and charged with manufacturing marijuana [48].

Kyllo filed a motion to suppress evidence stating that the use of the thermal imaging camera to obtain a warrant constituted an unlawful search. The district court denied the motion. On appeal, the Ninth Circuit Court of Appeals remanded the case for an evidentiary hearing. The district court affirmed the validity of the warrant and the defendant appealed again. The Ninth Circuit Court of Appeals initially reversed the ruling but eventually decided on affirming the district court's position. The Supreme Court decided this case was important enough that they granted certiorari and reviewed the case [48].

The Supreme Court concluded that when the government uses a device that is not in use by the general public to find details about a home that would have been unknown without a physical search, then the surveillance is a search and is unreasonable without a warrant [48].

3.3.12 Illinois v Caballes 543 U.S. 405 (2005)

The use of a drug-sniffing dog during a traffic stop does not violate the Fourth Amendment

An Illinois State Trooper stopped Roy Caballes for speeding on an interstate highway. A second trooper overhear the call to dispatch and proceeded to the scene. The second trooper

was a member of the Illinois State Police Drug Interdiction Team and had a narcotics canine with him. Upon arriving the officer walked his dog around the car while the first trooper was writing a warning ticket. The dog alerted to the trunk. Upon inspection, the officers found marijuana and arrested Caballes. He was convicted of a narcotics offense [49].

At the trial, the motion to suppress the evidence was denied. The judge determined that the officers had not unnecessarily prolonged the stop to do the search and that the alert by the drug dog provided probable cause for searching the trunk of the car. The court of appeals agreed but the Illinois Supreme Court reversed the decision because the dog was used when there was no evidence of drug activity. The court concluded that the use of the dog expanded the scope of the traffic stop illegally [49].

The Supreme Court granted certiorari to answer whether the use of a drug-detection dog during a traffic stop required reasonable suspicion. The Court decided that it wasn't. There is distinction between the legitimate expectation of privacy regarding information about legal activities and a criminal's expectation concerning the finding of contraband. The use of the dog as a method to only expose illegal contraband while maintaining privacy of all legal activities is not a violation of the Fourth Amendment. The Supreme Court vacated the decision of the Illinois Supreme Court [49].

3.3.13 United States v Brooks 427 F.3d 1246 (10th Circuit 2005)

A warrant should not limit the way in which a computer search is conducted

In August 2003, county law enforcement officials responded to a call that an unattended child had been left at the residence of Brent Ray Brooks. When the officers arrived they detected the scent of marijuana and obtained a search warrant authorizing the search of his residence for drug paraphernalia. Local law enforcement executed the search the next day and during a search of the garbage found significant amounts of what appeared to be printed child pornography images. Officers obtained a second warrant to search Brooks' home, including computer equipment, for child pornography. They also contacted the Federal Bureau of Investigation (FBI) for assistance with conducting the investigation [50].

When they executed the warrant, a federal law enforcement agent asked for Brooks' consent to search his computer for image files related to child pornography. Brooks was told

that an automated tool would be used to search for only images and that they would be displayed in a thumbnail format. Brooks' consented and signed a written consent statement that stated he authorized a complete search including a pre-search for child pornography of his computer tower. The pre-search automated tool did not work so the agent conducted a manual file search for images on the computer. After finding several images of suspected child pornography, officers shut the computer down and seized it. They obtained a third warrant to search the computer, two other additional computers, and several compact discs (CD) and diskettes. The forensic analysis was conducted at a police laboratory [50].

Brooks was charged with possession of child pornography and filed a motion to suppress the evidence found on his computer arguing that the officers exceeded the scope of his consent when they searched his computer by other means than what was explained to him. It also stated that the third warrant for the search of the computer did not sufficiently specify the search methodology making it a general search in violation of the Fourth Amendment. The district court denied his motion, and upon entry of a guilty plea, Brooks appealed the decision [50].

The Tenth Circuit Court of Appeals examined the two matters: whether the scope of the search was exceeded and whether the third warrant essentially authorized a general search because it did not specify the means through which the search would be conducted. The court found that the scope of the search was not exceeded given what had been granted in Brooks' written consent. While the original method did not work, the search conducted by the agent manually accomplished exactly the same task. While the means may have been different, the search was conducted correctly and within the scope of the consent. On the second point, the court points out that the method in which a search is conducted has never been required as part of a warrant. In the court's opinion, the search for evidence on computers is as much art as science and it would be difficult to place restrictions on the method of the search given the dynamic nature of computer forensics. In this case, regardless of the methodology used, officials sought out warrants at all appropriate steps and always limited the search for data pertaining to what was specified in the warrant [50].

3.3.14 United States v Ickes 393 F.3d 501 (4th Circuit 2005)

Search of a person's computer and digital media is permissible under the border search exception

John Ickes had his vehicle searched while attempting a crossing of the U.S.–Canada border near Detroit, Michigan. In addition to drug paraphernalia and photo albums of suspected child pornography found in his van, a search of his computer and 75 disks he had in the vehicle also produced evidence of child pornography. He was charged, among other things, with transporting child pornography. Prior to the trial, he filed a motion to suppress the evidence found on his computer on the basis that Congress had not authorized the search of his computer and disks as part of a border search and such a search is therefore unconstitutional [51].

Both the district court which convicted him and the Fourth Circuit Court of Appeals disagreed with him. Referencing sections of the United States Code that clearly empowers customs and border agents to search persons or property entering the country at anytime. They stated that the disks and computer were clearly classified as items being transported across the border and were subject to being searched. The court also pointed out that, while the types of searches done at the border may require a warrant at other locations, the border search exception has been in place for as long as the Fourth Amendment has been. Congress recognized early in the creation of the nation that it is a vital security and sovereignty issue to be able to screen and scrutinize the people and property entering the country [51].

3.3.15 United States v Grimmett 439 F.3d 1263 (10th Circuit 2006)

Computer search may be as extensive as reasonable required to locate items described in warrant

A detective with the Shawnee County Kansas Sheriff's Office applied for a warrant for the search of the residence of Stephen Grimmett and any computer equipment contained within it. The affidavit for the warrant was based upon testimony from a confidential informant who stated that she had seen child pornography within his residence. When the Sheriff's Office executed the warrant they seized a hard drive from a computer and viewed its contents on a laptop they brought with them. After the initial search, the lead detective

asked Immigrations and Customs Enforcement (ICE) to conduct forensic examination of the computer [52].

After making a sector-for-sector copy, the agent examined all the directories and folders. He testified to opening every folder but not every file, focusing only on those likely to contain child pornography such as images and movies. The examination found 1,500 images and 142 movies of child pornography. Grimmett was charged with possession and production of child pornography. The district court denied his motion to suppress evidence and he was found guilty [52].

Grimmett appealed the decision of the district court to deny the motion to suppress evidence. The Tenth Circuit Court of Appeals reviewed the district court's reasons for their decision: no second warrant was needed to conduct the search by the ICE agent and the search was not an impermissible general search. The court of appeals found no error with the district court's decision. The court noted that the warrant specifically authorized offsite examination of the computers and that two warrants are not required to seize and then search a computer when the search is for the same reason as the original seizure. On the second point, the court noted that while officers must be clear as to what they search for, a computer search may be as extensive as reasonably required to find items described in the warrant. The court found no evidence of the ICE agent rummaging through files. He limited his search to what he believed would contain child pornography. The court of appeals affirmed the decision of the district court [52].

3.3.16 United States v Hill 459 F.3d 966 (9th Circuit 2006)

A board seizure of digital media may be authorized as part of a warrant and the search methods to find evidence on that media need not be specified

Justin Barrett Hill was having his computer repaired when a technician discovered what appeared to be child pornography on the computer. The technician contacted the police who obtained a warrant based on a sworn affidavit stating that the technician saw two images of child pornography on the computer. The computer and all media related to it was seized. During forensic examination, more images were found particularity on some of the zip disks seized [53].

The defendant challenged that the warrant was overboard because it allowed seizure of all media without checking whether the media contained relevant material and because the search of the media had no limitations placed upon it. The Ninth Circuit Court of Appeals determined that the seizure of all materials was reasonable. It is unreasonable for police to make that determination on-scene when executing the search because technology is varied and complex. They also held that the search methods need not be specified because limiting police would make hiding evidence of a crime easy. Police must examine the files just as they would examine a bag containing a white powdery substance labeled flour [53].

3.3.17 United States v Heckenkamp 482 F.3d 1142 (9th Circuit 2007)

People have a reasonable expectation of privacy with respect to computers and digital media

Jerome Heckenkamp was allegedly using his personal computer on a university network for illegal activities including breaking into the school's email server and gaining unauthorized access to a Qualcomm Corporation network. While law enforcement officials were still investigating the break into the corporation's network, a university system administrator examined logs and determined the source of the illegal activity on the network. Realizing that the school's network was still at risk, the administrator, along with university police, took action to end Heckenkamp's access to the network. He was later charged with several offenses involving accessing a protected computer system without authorization. Heckenkamp attempted to file a motion to suppress evidence of his activities gained from the examination of his computer saying the search by the system administrator and police violated the Fourth Amendment. The district court denied the motion [54].

The Ninth Circuit Court of Appeals upheld the decision by the district court. This case is not important for that reason though. The court agreed with Heckenkamp that he did have a reasonable expectation of privacy regarding his personal computer that was legitimate and objectively reasonable. However, due to the special circumstances of the case, an exception to the warrant rule applied. The ruling in this case serves as one of the bases for computers being treated as a closed container from a search perspective, meaning that in most cases, the search or seizure of a computer would require a warrant unless an exception to the warrant rule applied [54].

3.3.18 United States v Arnold 533 F.3d 1003 (9th Circuit 2008)

Search of computers and electronic devices without reasonable suspicion is permitted during border searches

Michael Arnold arrived at Los Angeles International Airport after a trip to the Philippines and was searched by Customs and Border Protection (CBP) agents as he attempted to immigrate back into the country. The search of his luggage produced his laptop and several external storage media including an external hard drive and compact discs. The agents had Arnold start his laptop and proceeded to examine folders on the desktop which contained pictures that appeared to be of child pornography. After the initial discovery, special agents from Immigrations and Customs Enforcement interrogated Arnold. He was released but his laptop was seized. A warrant was obtained and several weeks later Arnold was charged with transporting child pornography across international lines and possession of child pornography. Arnold filed a motion to suppress the evidence found by the agents stating the search had violated his Fourth Amendment protections because the agents did not have reasonable suspicion to search his computer [55].

The motion to suppress was initially granted by the United States District Court of California, but the motion was reversed by the Ninth Circuit Court of Appeals. The court sited the Supreme Court that had previously ruled that reasonable suspicion for border searches of property is only required when there will only be exceptional damage to property or the search will be particularly intrusive. The search of the laptop did not meet either of these requirements, so the search conduced by the CBP and ICE agents was legal. A personal computer falls under the same category as a briefcase or luggage and can be examined without reasonable suspicion [55].

3.3.19 United States v Cartier 543 F.3d 442 (8th Circuit 2008)

Hash values can be used to establish probable cause, in addition, a search warrant need not specify a search strategy in order to protect private data

The Spanish Guardia Civil Computer Crime Unit (SGCCCU) conducted a sting operation in which they logged into a peer-to-peer (P2P) file sharing network and made available images of know child pornography seized in other investigations. Using hash values, they kept track of which files were downloaded by which Internet Protocol (IP) addresses. One

of the addresses was tracked to an ISP in North Dakota. The SGCCCU contacted the FBI and turned over the evidence collected. The FBI tracked the IP address to an address belonging to Steven Cartier and obtained a warrant to search his home. The warrant resulted in the seizure of 15 hard drives and thumbdrives and hundreds of compact discs and videos. Over 1,000,000 images and 4,000 videos of child pornography were found. The district court denied his motion to suppress evidence and he was found guilty of all indictments filed against him. He then appealed the decision to the Eighth Circuit Court of Appeals [56].

Cartier's appeal stated that the district court erred in denying the motion to suppress evidence because the agent failed to establish probable cause, the warrant was overly broad, and it did not articulate a search strategy. His argument was that probable cause was not established because the affidavit relied upon hash values of digital files and no agent had seen the pictures prior to the warrant search. The court of appeals reviewed the testimony about hash values given to the district court. The defense expert testified that hash values could collide meaning that two files could produce the same value which introduces doubt as to the actual contents of the file. The government expert testified that two dissimilar files would never have the same hash value and that collisions had only been found in controlled settings. The court ruled that, while the images had not been viewed, the hash values in combination with the other circumstances combined to produce fair probability and establish probable cause [56].

The Court also addressed whether the search was overly broad. Cartier's argument was that the warrant did not describe a search strategy, making it invalid. The Court cited many other cases which reject the notion that a search warrant must specify how a search is going to be conducted even with digital media. The court also noted that Cartier failed to show how the government used any unrelated files or even if they did search the unrelated files. The court of appeals affirmed the district courts decision [56].

3.3.20 United States v Crist 627 F.Supp.2d 757 (DC Middle PA 2008)

Warrantless hash-based search exceeded scope of private search

Crist had been evicted from his apartment and his computer given to a friend of the landlord. The friend, while going through the computer, discovered several movie files that appeared to contain child pornography. He deleted them initially but then decided to call the police.

In the meantime, Crist had filed a police report that his computer had been stolen. The computer was turned over to the Pennsylvania Attorney General's Office for forensics. The lab, under the impression that consent of the owner had been given, did a hash-based search of the hard drive for child pornography. There were abundant matches but at no time was a warrant obtained. A year after the computer was turned over to the lab for examination, Crist was interviewed by agents from the FBI. The agents revealed they had discovered the images and Crist admitted to having the files on his computer. He was arrested and charged with knowingly receiving and possessing files containing child pornography. Crist filed a motion to suppress the evidence found on his computer stating it was an illegal search [57].

The motion was brought before the District Court of Middle Pennsylvania. The government attempted to argue on several fronts that the evidence should be allowed but the most pertinent argument they made was that certain forensic tools do not constitute a search because of the way they work. The court ruled that because the software revealed more to the police than the landlord's friend, a warrant was needed to search the hard drive and that the all evidence gathered as a result of the computer search be suppressed [57].

3.3.21 United States v Giberson 527 F.3d 882 (9th Circuit 2008)

Computers, like briefcases and cassette tapes, can be repositories for documents and records

Francis Giberson was stopped by North Las Vegas police for an expired vehicle license plate. The officer determined that his identification card was fake and that he had three outstanding arrest warrants. His excuse for the false identification, among other reasons, was to avoid child support payments. An agent for the United States Department of Health and Human Services began an investigation into the child support obligations and discovered that a Minnesota court had ordered Giberson to pay child support, that he was behind $108,000 and that, ironically, he had served as the Deputy Commissioner of the Minnesota Department of Human Services. The agent obtained a warrant to search Giberson's residence in Las Vegas for records of financial assets, identification cards, other aliases, and employment. When the warrant was executed a personal computer was located inside the residence. Hooked up to the computer was a printer with printouts of fake identification (ID) cards. The computer was seized and a digital image of the hard drive stored until a

second warrant was obtained to search it for evidence of identification fraud [58].

An analyst began a search of the digital image. The software used sorted files into their type and put them into separate folders. While looking through the images, the analysts discovered what appeared to be images of child pornography. The analysts stopped his search and immediately contacted the FBI. The agent told him to continue his search for evidence of items related to fake IDs but to print any child pornography images discovered incidentally. The FBI agent obtained a third warrant to allow searching of the computer for child pornography. During the execution of the third warrant over 700 images where located. Giberson was charged with receipt and possession of child pornography. Giberson filed an initial motion to suppress the evidence which was denied by the district court. He plead guilty and appealed the decision to deny the motion to suppress evidence [58].

The Ninth Circuit Court of Appeals reviewed the case. Giberson's appeal challenged the seizure of his computer as part of the first warrant and the discovery of child pornography as part of the second warrant. Giberson argued that the seizure of his computer exceeded the scope of the warrant. He stated that due to the large amount of data that can be stored on a computer, especially private data, that a computer should have its status elevated so that a search and seizure of it be specifically stated in the warrant. The court disagreed saying that an exception such as this cannot be based on a technology as technology changes to fast. They also added that the argument of privacy is invalid because the Supreme Court had already established that a higher standard than probable cause should not apply when dealing with privacy implications. The court stated that the seizure was reasonable in order to protect evidence and that the officer's actions were appropriate because they obtained a second specific warrant to search the computer [58].

The second argument from the appeal stated the evidence of child pornography should be suppressed because the government did not limit its search for relevant documents. Giberson argued the methods used were too general. The court rejected this argument for several reasons. There was no reasonable way to sort images into relevant and irrelevant files, especially since data can be intentionally obscured or hidden. It is not the court's place to determine the means by which the warrant is executed. In addition, unlike in *United States v Carey* (p. 35), the analyst continued his search only looking for evidence of ID fraud. The search for images was authorized under the warrant as images had been previously found

relevant the investigation. The actions by the analyst make the case significantly different from the *Carey* appeal. The court of appeals held the search reasonable and affirmed the district courts decision to deny the motion to suppress the evidence [58].

3.3.22 United States v Comprehensive Drug Testing, Inc 579 F.3d 989 (9th Circuit 2009)

New guidelines established for an investigation that will examine digital media

This case started as an investigation into a company suspected of providing steriods to professional baseball players. As part of the investigation, federal agents attempted to subpoena Comprehensive Drug Testing, Inc (CDT) for all the confidential drug testing records they had. The negotiations between the company, the government, and the players association failed so the investigators applied for a warrant to search CDT's offices for the records of ten players for whom they had probable cause. The warrant specified a fairly broad seizure of computer information due to the difficultly of retrieving electronically stored data. The magistrate judge allowed the seizure but placed restrictions and safeguards in an effort to protect the Fourth Amendment rights of anyone associated with the seized information. The execution of the warrant and subsequent search of the data was mishandled by the investigators and did not comply with the limitations imposed by the magistrate. A motion filed by CDT and the players moved for the return of the seized data which the government appealed to the Ninth Circuit Court of Appeals [59].

The court, referencing *United States v Tamura* (p. 28), stated that in the past the collection of massive amounts of data in order to search for a small amount of evidence was the exception to the rule and could be handled that way. In the electronic age, the exception has now become the rule. Data are commingled and extensive on even the most basic personal computer. The result of the appeal was a set of guidelines established by the court of appeals for the government to follow when seeking a warrant to examine digital media [59]:

- The investigators should waive reliance upon the plain view doctrine.
- Segregation and redaction must be done by specialized personnel or independent third party so that information outside of the warrant is not released.
- The risk of destruction of information as well as prior efforts to seize the information

must be documented.

- The search protocol must be designed to uncover only the information for which it has probable cause, and only that information may be released to case agents.
- The government must return or destroy non-responsive data.

The court of appeals called on the sound judgment of the magistrate to enforce these guidelines and find the proper balance between the protection of the rights of citizens and safety of the community [59].

3.3.23 United States v Mann 592 F.3d 779 (7th Circuit 2010)

A hash search for known illegal content exceeded the scope of a warrant but the plain view doctrine applied

Matthew Mann was a life guard who covertly installed a video camera in the woman's locker room at the pool he worked at. A female student of his discovered the camera, which included footage of him installing the camera, and turned it over to local police. A search warrant was granted to search Mann's residence for video tapes, CD's, digital media, computers, and their contents for images of woman in private places. The police seized his desktop, a laptop, and several external hard drives. He was charged with voyeurism the next day. Two months later the forensics examination of his computer began. The detective created a sector-for-sector image of the drive and used FTK to catalog all the images found. The software also produced a list of alerts of files that were flagged as being previously known. These files are typically child pornography images. At the suppression hearing, Mann argued that the officers had exceeded the scope of the search when they opened the files containing child pornography. The district court disagreed stating that the officer never abandoned his search for evidence of voyeurism and that while a few of the images were found outside the scope of the warrant, they would still be covered under plain view. Mann entered a conditional guilty plea and appealed [60].

Mann's appeal stated that the search of his computers exceeded the scope of the warrant and that the plain view doctrine did not apply in this case. He focused specifically on the use of FTK and the Known File Filter (KFF) feature. Given the nature of the search, the court found the use of FTK to be completely appropriate since it would find, index, and catalog all the images and put them into a viewable format. The court did have issue

with the use of the KFF feature. The investigator should have known the four files flagged would likely be matches to child pornography images and would be outside the scope of the search. They found that the scope of the search was exceeded during the investigators examination of these four files. The rest of the analysis uncovered many other images of child pornography. They were found to be admissible as the intent of the investigator's search was to find evidence of voyeurism and these images were found in plain view. The court rejected the notion of dismissing the plain view doctrine based solely on the fact that an investigation is examining digital media. It pointed out, however, that it would have been better if, after discovering child pornography, the investigator would have sought a warrant for a separate search [60].

3.3.24 United States v Miknevich 638 F.3d 178 (3rd Circuit 2011)

A file name and hash value can be used to establish probable cause for a warrant search

Stephen Miknevich had his home searched and computer seized as part of an investigation into a child pornography P2P network. He was arrested and charged with possession of child pornography. He filed a motion to suppress the evidence found on his computer stating that the warrant was issued without probable cause. The district court disagreed so Miknevich plead guilty, was convicted, and appealed the decision to deny the motion to the Third Circuit Court of Appeals [61].

The court determined that its role in this appeal was to determine if the magistrate judge who issued the warrant did have a substantial basis for concluding that probable cause existed based on the facts available to him at the time. Probable cause is established when, after viewing the circumstances, there is fair probability that contraband or evidence of a crime will be found at a place. In this case, those facts were based on the submitted affidavit which is summarized in the following paragraph [61].

A Delaware State Police detective was conducting an investigation into a P2P file sharing network that was suspected of being used to distribute child pornography images and movies. The officer ran a search of the network for known terms associated with child pornography. The list returned included file names, types, sizes, and SHA-1 values. He recognized one file and its SHA-1 value as having been child pornography. Indicating that he wished to download the video, the network returned a list of computers sharing the file.

Using packet capture software, the detective determined the IP address of the person with the file. He then turned his results over to an investigator in the Delaware Country Pennsylvania Internet Crimes Task Force, who filed for and received a court order for Comcast Cable Communications to supply the user information of the account assigned the IP address during the specific date and time [61].

There is controversy surrounding the affidavit as it was unclear if either officer actually downloaded and viewed the contents of the video. Miknevich's main argument is that a file name and hash value are not enough to establish probable cause. The court of appeals disagreed. They assumed, for arguments sake, that neither of the officers nor the magistrate judge viewed the contents of the file and still found that a judge could have drawn reasonable inference as to the contents of the file and establish probable cause based on a combination of the descriptive file name and the SHA-1 digest. The court specifically pointed out that the hash value is both relevant and very important as it serves as a digital fingerprint and due to the fact that file names are often very inconsistent with a files contents [61].

3.3.25 United States v Cotterman 709 F.3d 952 (9th Circuit 2013)

The border search exception does allow forensic analysis conducted off-site of electronic media with reasonable suspicion.

Howard Cotterman was stopped at the U.S.–Mexico border after a search in a database returned a hit for a fifteen-year-old conviction involving child molestation. The entry indicated that he was suspected of being involved in child sex tourism. Cotterman and his wife were referred to secondary inspection. A search of the vehicle revealed two laptop computers and three digital cameras. Inspection of the computers at the crossing revealed personal and family photographs and several password-protected files. While the search was being conducted, agents at ICE, who had made the database entry, were contacted for more information. ICE agents decided to interview the Cottermans personally and seized their laptops for forensic examination. Cotterman offered to unlock the password protected files but the agents declined, worried that evidence would be deleted. The Cotterman's were allowed to leave the border crossing several hours later after being interviewed, but their laptops and cameras were retained by the ICE agents [62].

The agents drove the laptops and cameras 170 miles to the ICE office in Tucson Arizona where they were turned over for forensic analysis. Initial examination found 75 images of child pornography in the unallocated space of the hard drive of Cotterman's laptop. After being contacted to unlock the password protected files, Cotterman fled the country to Australia. Analysts were eventually able to unlock the password protected files and discovered over 350 additional images some of which included Cotterman himself. He was indicted on several offenses related to child pornography and filed a motion to suppress the evidence gathered on the laptop. The magistrate judge had filed a finding classifying the forensic examination as an "extended border search" which would have required reasonable suspicion. The district judge agreed, determined that reasonable suspicion did not exist based solely on the database hit, and granted the motion to suppress [62].

The case was originally appealed to the Ninth Circuit Court of Appeals and a divided panel reversed the decision, stating that a border crossing laptop search was allowed off-site without reasonable suspicion when it remained in the custody of government agents. A rehearing on the matter was reordered by a majority of non-recused judges. After complete review, the court of appeals determined [62]:

- The initial search at the border crossing was completely legitimate. No suspicion is required to conduct a quick and unobtrusive search of laptops at border crossings.
- The follow on forensic analysis was not an "extended border search" as the laptop was never returned to Cotterman. Cotterman never regained an expectation to privacy in regards to the laptop as he would have if it cleared customs. The fact that it was taken off-site for a follow-on examination does not make it extended.
- Due to the intrusive nature of the examination, such as the recovery of deleted files and the decryption of password protected data, the forensic analysis of computers at border crossings requires reasonable suspicion regardless of where it is conducted.
- In this case, the totality of the circumstances established reasonable suspicion which justifies the forensic examination. The court noted that the presence of password protected files did contribute to this although just the presence of password protected or encrypted data alone would not be enough to establish reasonable suspicion.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
Previous Analysis

To date, there does not appear to be specific work discussing how new forensics techniques such as block-based hashing, random sampling, and similarity matching apply in the context of federal law and in particular the Fourth Amendment. This is not surprising given that many of these particular techniques have yet to be implemented by the forensic community at large. Several researches have examined how the laws governing search and seizure do apply to computer forensics in general, however. This chapter presents some of that literature.

4.1 Searches and Seizures in a Digital World

Kerr has examined the topic of the Fourth Amendment and the search and seizure of computer data extensively [30]. He specifically addresses how the Fourth Amendment should regulate the process that an investigator follows when attempting to retrieve evidence from a computer.

4.1.1 Differences Between a Physical Search and Computer Search

Searches of computers and physical locations are similar in that their purpose is to locate pertinent items that are hidden. There are four major differences between the search of a physical structure, such as a home, and search of a computer that necessitates the review of how the Fourth Amendment applies to the area of digital forensics:

- **The environment:** The Fourth Amendment's purpose is the protection of citizens and their property, especially their homes. The basic mechanism of searches envisioned by the Fourth Amendment is for the investigator to enter into a physical space, observe, and move property to make additional visual observation. The search of a computer is completely different. Data are accessed from storage devices, transmitted electronically, interpreted by the processor, and displayed for inspection.The investigator may be present or remote [30, p. 538].
- **The copying process:** When law enforcement officials conduct a physical search they typically do it at a location associated with the subject of the search and the

search is of property associated with the subject. With a computer search, the computer's storage may be searched directly by a program run on the computer by the investigator or an exact image of the digital media in question may be made. This copy, not the original media, may be analyzed in another location using government computers [30, p. 540].

- **The storage mechanism:** An important difference between physical and digital searches is the amount of storage available and the control of the content of what is stored. A location that is subject to a search, such as a home, can contain a lot of items but it is physically limited by the size of the structure. A computer is small in size but can contain the equivalent of a warehouse full of data. The owner or user of the location also typically controls what is being stored at a given location. Computers contain whatever data their users have stored on them but in addition they may contain other data downloaded by software without the user's knowledge. The user thus has little to no control over some of the data on the computer [30, p. 541].
- **The retrieval mechanism:** The search of a physical location is done by a specially trained team at a specified location looking for specific items. When the search is done, the officers leave. Computer searches typically require fewer people but significantly more time. The search can be done with reference to the file system or not. Thus, analysts can search for files that are hidden, deleted, or intentionally modified to be difficult to find. Indeed, the government analyst can recover data that is invisible to the computer's owner. Because of this, it is very difficult to put restrictions on the techniques and methods used along with a standard time line [30, p. 543].

Kerr then focuses on the two major steps of a digital forensics investigation: the acquisition of data and the reduction of data. An examination of relevant case studies points out some of the current policy and issues with the way these steps are currently treated. He then proposes a set of rules for how this process should be conducted.

4.1.2 Data Acquisition

The data acquisition step includes all efforts to gain access to a computer system and the collection of information to be searched. There are two aspects to this that need to be addressed: rules that govern looking through a computer and rules that cover the creation of

a digital image of digital media [30, p. 547]. The range of activities associated with looking through a computer are pretty broad but in general it deals with any time a law enforcement official uses the operating system to view several files on a computer. Possible examples include browsing a computer as part of a search of a home or turning on and examining a computer that a private citizen turns over to officials because he or she saw evidence of illegal activity on it. Many courts have ruled, and Kerr agrees, that the accessing of the contents of a computer or other digital media constitutes a search which requires consent, a warrant, or an authorized exception to the warrant rule [30, p. 550].

Two things need to be addressed more throughly: when the search begins and the scope of the search. Rather than deal with the technical aspects of when the data is actually accessed, Kerr offers a very simple solution: the search begins when data or information about the data are exposed to possible human observation. He calls this an "exposure based approach" [30, p. 548].

A similar approach can be applied to the scope of the search. Current precedent is inconsistent: sometimes a computer is treated like a container in which the thousands of files are treated like individual containers, sometimes just the physical device is treated as a container so once permission is obtained to inspect it, all data inside is subject to review. Kerr once again proposes using an exposed information approach: the scope of the search would be defined as whatever information appears on the output device [30, p. 556]. Examining unexposed data would constitute a distinct search of that data which may be authorized depending on circumstances.

The other aspect of data acquisition is the creation of digital images and how those images are treated. In most cases, not involving digital media, the creation of a copy does not represent either a search or a seizure. Expanding this to cover data would legally allow the government to copy and store data with no restrictions. This is not what typically occurs in most criminal cases however: what usually happens is the computer is considered seized while the imaging process is being conducted, the act of which must meet Fourth Amendment requirements [30, p. 561]. In addition, proper authorization is still required to search the data. Kerr would like to see this more formally implemented to ensure protection of privacy.

The law does not currently codify how the digital images are treated from an evidence perspective: as an original or as data stored on a separate machine. Some police departments treat them as property of the court. In practice, digital images are treated as if they are originals and strict controls are put in place to ensure their integrity. Again, Kerr believes this should be established in law [30, p. 562].

4.1.3 Data Reduction

Data reduction refers to the search through a digital image for evidence related to a crime [30, p. 565]. Kerr addresses the kinds of warrant searches that should be reasonable and unreasonable and the rules that should be put in place to regulate police actions both before and after the discovery of evidence.

Even if a warrant is very specific in the type of information being searched for, due to the implementation of technology and the actions of users, a very broad search may be needed to find relevant information. During physical searches a balance must exist that allows police to be able to act upon evidence of criminal activity in the course of any search but at the same time protect the privacy of individuals from discriminatory and pretextual searches. The plain view doctrine [30, p. 568] is the legal compromise that attempts to accomplish this. It permits the police to seize evidence of crime discovered during a valid search even when it is unrelated to what is observed. For example, if police are summoned to a house for a domestic violence case they may seize drug paraphernalia if they see it on a table in plain view.

Kerr states searches of digital media are trending to become general searches and become invasive for several reasons [30, p. 569]:

- The way data is physically and logically laid out on the media.
- The use of computers in the lives of people continues to increase, so more personal data is being stored often with the user not realizing it.
- For every method developed to find data on digital media, users attempt to find more complex ways to obscure or hide it for both legal and illegal reasons.

For these reasons, the forensic analyst must often employ techniques that involve exposing a great deal of information in order to be confident that the relevant evidence has been

found. This poses the risk of exposing information not pertinent to the investigation. To protect the privacy rights of people, the judiciary can establish limitations before the search is executed, called *ex ante*, or after all the information is exposed and the admissibility of evidence is later determined, called *ex post*. Kerr argues that attempting to restrict the search beforehand is not viable based on the unpredictability of the forensic process [30, p. 572]. The judiciary is poorly equipped to determine what protocols are applicable ahead of time that will protect privacy while not hindering the analyst's search for relevant data.

Kerr proposes that restrictions be made after the search is complete but, in order to protect privacy, the plain view doctrine would need to be reevaluated specifically when a search of a computer is conducted. The plain view doctrine is what allows for evidence of a different crime to be admissible when it is found as part of another valid search. Evidence of illegal activity in plain view of the law enforcement officer can serve as probable cause for subsequent searches. In the search for data on a computer, the plain view doctrine in combination with thorough forensics techniques could lead to a general search. Instead, Kerr argues that the plain view doctrine be abolished which would allow the analysis access to all forensic methods without restriction but only data pertinent to the search would be allowable as evidence unless inevitable discovery applied [30, p. 577].

4.2 Fourth Amendment Search and the Power of the Hash

Salgado built upon the exposure framework introduced by Kerr and addressed how hash algorithms are used in digital forensics in that context [63]. His work briefly explores the properties of hash algorithms and their uses at the file level and above in forensics: ensuring data integrity, searching for known content, and excluding known content. He then addresses specifically the use of hashing for integrity validation and data reduction in the context of the Fourth Amendment.

4.2.1 Integrity Validation

Kerr proposed that a Fourth Amendment search occurs anytime data is exposed to human observation. However, when digital media is imaged, it is common to calculate a cryptographic hash value to allow integrity verification at a point later in time. Should exposure of the data to the hash algorithm be treated as a search because the calculation of the hash value involved every bit, or should it not be considered a search because nothing knowable

about the original data is exposed? Salgado proposes that hashing does not constitute a search because the value is derived, the actual degree of intrusion is minimal, and nothing about the data is exposed [63, p. 42].

4.2.2 Data Reduction and Exposure

Hash algorithms are also used to find known data. When a hash value of a file extracted from a digital image matches a hash value from a set of known data it can be used to exclude the data, data reduction, or highlight its existence, data exposure [63, p. 43]. When hash algorithms are used to reduce the amount of data to be searched, the examination not only occurs more quickly but the search is also less intrusive as non-pertinent data is not exposed to human eyes. Under Kerr's framework, exposing data using hash-based searches would constitute a Fourth Amendment search. Salagado extends this one step and addresses searching the data for illegal content without a warrant or exception to the warrant. His argument centers on the precedent set by *United States v Jacobsen* (p. 29) and *Illinois v Caballes* (p. 37) which determined that the use of certain methods and techniques were allowed when the test would only expose known contraband. The use of chemical field tests to test for narcotics or drug sniffing dogs are the relevant examples in these two cases. In the case of digital forensics, an argument can be made that current precedence would allow that a hash-based search of known illegal content be permissible during any search regardless of scope of the search. This would be justified because the results of the search would only show known illegal content and the invasion of privacy would be minimal [63, p. 44].

4.3 Judicial Confusion and the Digital Drug Dog Sniff: Pragmatic Solutions Permitting Warrantless Hashing of Known Illegal Files

Burrows specifically addressed searching for know illegal content using file-based hash searches. Her work [64] examines how the judiciary has been inconsistent in its rulings regarding the use of hash searches. She believes that the use hashing to search should not constitute a Fourth Amendment search but states that the judiciary is likely never going to accept this. Burrows then proposes three rules that would permit file-based hash searches

without a warrant while maintaining the privacy interests of the general public.

4.3.1 The Digital Examination

Burrows discusses how a typical computer forensics investigation takes place. She describes the process as exacting as all steps must be documented, any items seized inventoried, and careful steps taken during both on-site and laboratory analysis to ensure any original media or copies thereof are not compromised. She states that forensic examination of digital media can take place at the physical and logical levels. Evidence can be found using many different techniques including file carving using file signatures and searching for data using the file system. Using the hash values of files to search for known illegal content is a means to speed up a search for specific data. She details how hashing algorithms work and how they are used in the forensics process: ensuring data integrity and searching. Hash value matching functions are contained in many of the standard forensics analysis softwares including FTK and EnCase [64, p. 259].

4.3.2 The Fourth Amendment and Digital Evidence

Burrows states that the Fourth Amendment was designed to ensure that general warrants, such as those used by the British government in Colonial times to search homes, would not be allowed in the United States. Any search of a citizen's person or property by the government must have a warrant issued with probable cause, particularity, and reasonableness. The concern is that using hashing allows for a specific search to degrade to a general search because so much data, which may have a privacy expectation, is commingled and has an algorithm applied to it [64, p. 264].

4.3.3 Judicial Inconsistency

The judiciary has been inconsistent in its rulings regarding file-based hash search. Burrows sites multiple cases covering the use of hashing to find evidence covering the gambit from consent searches to warrant searches that discover evidence of another crime. In some cases the courts find the search and use of hash value matching reasonable; in other cases they do not. She then presents an analysis as to what has caused this apparent split and makes the argument that much of the confusion stems from how the methods work and the courts lack of understanding as to what the technologies do [64, p. 270].

4.3.4 Hashing is Not a Search But Courts Hesitant to Agree

Burrows makes the case that a hash-based search should not be considered a Fourth Amendment search. She references the arguments that Saladago made and the one-way property of cryptographic hash functions which makes knowing anything about the original data based on just the digest not possible. She also argues that hashing can be used to make investigations less intrusive because they can be used to either exclude content or search for very specific content. She also points out that the use of hashing without a warrant is not without precedent as hashes were used in *United States v Miknevich* (p. 49) to establish probable cause [64, p. 276].

Burrows believes that, despite sound arguments that hashing is not a Fourth Amendment search, the judiciary will be very hesitant to allow hash-based searches in the way that field tests and drug sniffing dogs are allowed to be used to search for illegal items without a warrant. There are several factors that come into play [64, p. 280]:

- The methods and techniques are not as well understood by both the public and the courts as it is in the case with a dog.
- The search often requires examination off-site.
- A drug sniffing dog would not be suspicious of having ulterior motives whereas an analyst may have suspicions aroused when using a tool and database created by humans and trusted to be run correctly by humans.

4.3.5 Solutions That Allow Hashing Without A Warrant

Burrows believes that the courts will come to see that hash-based searches do not qualify as Fourth Amendment searches and that time and compromise may allow for hash-based tools to be trusted and used without a warrant. In the interim, she suggests three solutions that will allow for warrantless hashing while keeping privacy issues in mind [64, p. 281]:

- **Create a suppression rule for hashing tools identifying known illegal files:** A case log, which documents all steps the analyst took during the course of an investigation, should be required as part of any evidence submitted as admissible to the court. These logs allow for the search process to be repeatable and would give the courts a means to suppress evidence if its case log shows that the actions of the analyst exceeded the score of a search [64, p. 281].

- **Perform hash analysis on-site:** In an effort to make a hash-based search similar to using a drug sniffing dog, the hash analysis could be performed on-site. This does come with risks and limitations, as on-site analysis is not conducted in a controlled environment and thus could damage evidence. It requires equipment to run the hashing and search at the scene of the investigation and depending on the size of the data, could require lots of time. The hope is that seeing the methods employed correctly would allow the individual and the public to see that privacy is not being violated [64, p. 284].
- **Demonstrate hashing in court:** Showing how the methods and techniques work in court would help both judges and juries understand what occurred during the forensic analysis and would educate them as to what occurs when a hash-based search for known illegal content occurs [64, p. 288].

4.4 The Physical Computer and the Fourth Amendment

Goldfoot examined how the Fourth Amendment applies to investigations involving computers [65]. He addresses the two most common ways to view digital media:the internal view as a container of sub-containers, or the external view as a physical object. His work examines both perspectives in detail including arguments for and against and then proposes that the physical object perspective is the proper view.

4.4.1 The Container of Sub-containers

The container of sub-containers perspective views digital media as a collection of individual groups of data that all are their own container. Each container requires justification for its examination. A container does not just exist at the file level. It can be above it, at the folder level for example, or below it as an individual line in a spreadsheet or a block of data of a file. The storage medium is treated like a collection of information only some of which may be used as part of the investigation. This is the most popular view among the courts and the legal profession at large [65, p. 118].

To accommodate this view of digital media, it is necessary to translate search and seizure law into something more logical and virtual. This is because almost all the laws, ranging from the Constitution to the U.S.C. primarily address the search and seizure of a physical

thing or location by government agents. Goldfoot contends that attempting to make this translation presents several issues [65, p. 123]:

- **Drawing sub-containers:** The goal of this perspective is to place barriers in place that regulate access to information so that the investigator does not look at too much information. In the digital world, these barriers are absent due to the way in which a computer is implemented. The normal division tends to be at the file level but Goldfoot points out that modern techniques do not just focus on whole files found on digital media. It also include the examination of Random Access Memory (RAM) and sub-file fragments. Implementing the container could be done internal to a file or at the sub-file level as well, but presents issues concerning what to redact and what not to in the course of an analysis. The container of sub-containers perspective is difficult to implement because digital media does not lend itself to being divided [65, p. 125].
- **Reasonable searches:** The laws governing the procedures for search and seizure are designed to protect the population from unreasonable searches and seizures. The requirements that government agents must meet prior to executing a search provide such protection. These rules do not translate well with the internal view of digital media. It is easy to determine when a place has been searched and items seized from it in order to serve as evidence. What a seizure is and when the search occurs has not been clearly defined with digital media [65, p. 131].
- **Regulating the forensic examination:** The container of sub-containers view struggled to properly regulate a forensic investigation. When the investigator leaves the location of a physical search with a sector-for-sector copy of a hard drive, the agent has all the data with him. The simple fact is that modern forensics requires sifting through large amounts of data to find the needle in a haystack. What ends up occurring is one of two extremes where either all the data ends up being examined anyway, making the point of having sub-containers mute, or the pertinent data ends up being suppressed because the search for it violated the protected status of the containers [65, p. 136].

4.4.2 The Physical Object

The physical object perspective treats digital media like any other object subject to search and seizure. With this view, once an investigator comes to have legal possession of the media in question, that person would be authorized to conduct whatever forensic procedures were necessary to find the relevant information on it. Digital media would be treated the same a blood found at a crime scene or an article of clothing with DNA evidence on it. Testing can be done on these items without specific requirements set by the judiciary. A physical view of digital media means that there is no need to translate search and seizure laws. A physical premise is searched and digital media seized as part of the execution of a search warrant. Information found on the media are now facts learned during the investigation [65, p. 149].

The primary argument against this approach is that the focus of the search is on the information the media contains, not the physical object itself. Goldfoot points out that this is true with any piece of physical evidence. It is the object combined with the analysis that give meaning to physical evidence during a trial. The results are then explained by an expert witness as required to the court or the jury in a case [65, p. 153].

4.4.3 The Debate on Which View to Adopt

Goldfoot argues that both methods lead to the same result: in the course of a forensic analysis, an examiner may end up viewing all the pieces of data. With the physical view, this can be done directly. With the container view, it is done through the haystack problem and the plain view doctrine. The container of sub-containers perspective would require special rules be implemented much as was done when the Supreme Court determined wiretaps were a Fourth Amendment search. Data is commingled and vast amounts of it can be stored on digital media. More and more of it is private as society continues to expand its use of computers. Its exposure with no controls grants law enforcement a lot of authority. The fundamental question is whether searching a computer is more like the search of someone's home, a large warehouse facility, or some other physical entity [65, p. 160].

4.5 Constitutionality of Cell Phone Searches Incident to an Arrest

As this thesis is being written, the Supreme Court is set to decide on the constitutionality of a search of the data stored on a cell phone by government agents incident to an arrest. It has granted certiorari regarding two cases summarized below.

4.5.1 Riley v California, No. 13-132 U.S. (2014)

David Riley was pulled over by the San Diego Police Department for driving with expired license plates. Upon confronting Riley, the officer learned that his drivers license was expired and impounded his automobile. An inspection of the vehicle was conducted at the police impound in accordance with department policy. Two firearms were discovered under the hood of the car. Riley was arrested for carrying concealed and loaded weapons. Officers seized Riley's "smart" cell phone while searching him incident to his arrest [66].

Two searches of the cell phone occurred while Riley was in police custody. Officers looked through the text messages on the phone and also examined the photo and video gallery. From the information gathered, officers deduced that Riley was likely a gang member and involved in a drive-by shooting involving a red car that Riley owned. This combined with the ballistics of the seized firearms lead to Riley being charged with shooting at an occupied vehicle, assault with a firearm, and attempted murder [66].

There were two jury trials as the first one resulted in a hung jury. In both cases, the motions to suppress the evidence found on the phone were denied. Riley was found guilty of all three counts in the second trial and received a higher sentence because the shootings were gang related. The California Court of Appeals heard Riley's appeal and confirmed the decision of the district courts based on a recent decision by the California Supreme Court to allow exploratory searches without a warrant of a person's cell phone if the phone is discovered incident to an arrest. Riley appealed to the California Supreme Court which denied review [66].

4.5.2 United States v Wurie, No. 13-212 U.S. (2014)

A Boston police officer observed an apparent drug sale out of the car of Brima Wurie. After the sale was complete, the officer confronted the buyer and discovered two bags of

crack cocaine in his pocket. The buyer told the officer that Wurie, the driver of the car, had just sold him the drugs. Officers following Wurie arrested him, read him the Miranda warnings, and took him to a police station. They seized among other things, two cell phones and over one thousand dollars in cash. Officers noted that one of the phones, a flip type, kept receiving a phone call from a location labeled as "my house" on the front screen. Officers eventually opened the phone, noticed a picture of a woman holding a baby as the background image, and navigated to the call logs where they retrieved the phone number associated with "my house" [67].

Officers conducted a database search for the house and found it associated with a home in South Boston. Suspecting that there would be a drug cache at the location, officers drove to the residence, confirmed Wurie's name on a mailbox, and observed through a window a woman whose description matched the picture on the phone. The officers obtained and executed a search warrant on the premise where they seized crack cocaine, marijuana, a firearm, and ammunition [67].

Wurie was charged with felony possession of a firearm, distributing crack cocaine, and possession of crack cocaine with intent to distribute. Wurie filed a motion to suppress evidence found from the search of the apartment stating that it was fruit of an unconstitutional search of his phone. The district court denied the motion stating that the search incident to arrest exception allowed for a search of the phone without a warrant. He was found guilt on all three counts [67].

On appeal, the First Circuit Court of Appeals reversed the decision, vacating the first and third convictions, stating flat out that the incident to arrest exception does not authorize a search of data or a cell phone seized from a person under arrest. The court followed this decision with an inquiry as to whether a warrantless search of data on a cell phone can ever be justified particularly with regards to the preservation of evidence. They found that the government's argument that a phone could be remote wiped justified an immediate search was insufficient to satisfy the Fourth Amendment [67].

4.5.3 What is Being Decided

The question before the Supreme Court is whether a search incident to an arrest exception to the warrant requirement authorizes government agents to search the data of a personal

electronic device found on or in control of the suspect when the individual has been legally arrested.

Those who believe the exception does not apply to electronic devices are concerned about the large amounts of sensitive information stored on them. They cite three reasons why a warrantless search of the phone is unconstitutional [68]:

- When the exception was envisioned it was during a time when the scope of the search would be limited by the amount of physical material carried by the person. Allowing the search of all data on electronics like phones now translates into rummaging through large amounts of data which is what the writers of the Fourth Amendment sought to avoid.
- The justifications for the search incident to arrest exception were based upon establishing no threat to the officer and to keep the suspect from destroying evidence. A seizure of the device and visual inspection will allow for both those objectives to be meet. If there is a risk of remote deletion then there are preventative actions the arresting officer could take to preserve the information while not searching it.
- A rule requiring a warrant before searching digitally stored information is clear and easy to implement.

Those who believe the exception does apply to electronic devices point to the popularity of these devices and state that if someone is worth arresting then there is a high likelihood that the suspects phone or other device contain evidence of criminal activity. They cite several reasons why the search is constitutional [67]:

- Past precedence supports the search of the suspect and the discovery and seizure of evidence of criminal activity. This same standard should apply to portable electronic devices.
- Even if it were appropriate to create specific exceptions exempting certain items from search there is no sound justification why portable electronics should be on the list when they are very likely to contain evidence pertinent to the arrest.
- The search incident to arrest limits the search to only the information stored on the phone by the very definition of the exception.

CHAPTER 5:
Hypothetical Scenarios

This chapter presents three scenarios and an analysis of how current federal law applies to the use of new digital forensics techniques. Hypothetical scenarios are used because they offer complete knowledge of each case, can be tailored to focus on the three types of specific searches addressed, and demonstrate the techniques currently being researched. It is important to note that federal courts in the United States never make determinations using hypothetical scenarios, instead they wait for issues to be presented in court with the background and context that an actual case provides.

The three scenarios involve three different types of searches and showcase three different forensics techniques. The three types of searches covered are consent searches, warrant searches, and border crossing searches. The forensic techniques showcased by these scenarios are block-based hash searches, random sampling, and similarity matching.

5.1 Consent Search of Vehicle Leads to Discovery of Cell Phone

Sector hashes can be used to find traces (1 to 100 blocks) of known content that would otherwise be missed, and to perform rapid analysis using random sampling when large amounts ($>$ 30 megabytes (MB)) of known content are suspected of being present. The purpose of the first hypothetical scenario is to explore the discovery of trace evidence. This scenario also examines how law enforcement officials could possibly exceed the scope of a search by using consent (which might be withdrawn at a later time) to collect block hashes for later analysis.

5.1.1 Scenario

An adolescent had been reported missing. Several weeks later, law enforcement officials obtain a cell phone video that portrays the adolescent's murder. A witness reported seeing the victim entering a vehicle on the day of the disappearance. Law enforcement officials conduct a Department of Motor Vehicles

(DMV) records search for vehicles matching the description and one of the matches is to a man named Dana. Officials question Dana at his home as to his whereabouts on the day of the abduction and ask for his consent to search his vehicle. Dana is cooperative and allows the search. The officers conduct a thorough inspection of the vehicle. While searching, investigators discover a smart phone with an SD card in the door's side pocket. They perform an inspection of the phone and discover that the phone is protected by a Personal Identification Number (PIN). They then remove the SD card and make a sector hash image of the SD card. Unlike a traditional byte-for-byte copy, the sector hash image contains only the hash of each sector on the digital media, and not the actual data. At the conclusion of their search, the team packs up their equipment, puts the SD card back in the phone, puts the phone back in the car, thanks Dana for his cooperation and departs.

The sector hash image is turned over to a digital forensics lab. The analyst is told the image was made as part of a consent search regarding the kidnapping and murder of a teenager. The analyst conducts a search for matches of the sector hashes from the SD card to hash values of blocks from the video that the police obtained. Ten sector hashes from the cell phone match block hashes from the video. The hashes are from noncontiguous blocks and the blocks do not form the whole file. Had just a file-based hash search been conducted, the matches would have been missed.

5.1.2 Analysis

Consent searches are unique in that the burden of proof that the person or thing being searched was consented to lies with the investigator. The person subject to the search has the right to withdraw consent at any time. Dana consents to the search of his vehicle on the basis that the investigators are looking for evidence related to a kidnapping and murder and presumably withdraws that consent when the officers leave.

The legality of making a sector hash image during a consent search

While the phone was found as part of the consent search, it is not clear whether specific consent is required to make a sector hash image of the phone's SD card. In *United States*

v Heckenkamp (p. 42) the court found that computers and other electronic devices are to be treated like closed containers and that people have a reasonable expectation of privacy regarding them. As shown in *Florida v Jimeno* (p. 32), a closed container in a vehicle can be searched as long as it is reasonable to assume the container may contain information pertinent to the search. The officers have reason to believe evidence may be on the phone as they have cell phone footage showing the murder.

A key difference here is what the agents did with the phone. Typical cell phone searches usually include inspecting the call logs, address books, text messages, and examining images. In this case, the agents performed a sector hash of the storage. Using the exposed information framework set forward by Kerr and Salagado, computing the hashes of all the sectors exposes nothing about the actual data itself to human eyes. A collection of pseudo-random characters means nothing when visually inspected. If that is the case, consent would not be required to make the sector hash image and keep it. The making of the sector hash image could also be considered a seizure but the government is often authorized to seize data and hold it until authorization for a search is made such as in *United States v Tamura* (p. 28).

On the other hand, a case can be made that making the sector hash image was an illegal search. The decision in *Kyllo v United States* (p. 37) demonstrated that the use of new technologies which do not discriminate in what they report, produce, or show are often considered searches under the Fourth Amendment and require a warrant. While the result of the sector hash image is a collection of pseudo-random characters and numbers, those hash values represent all the data on the cell phone including pictures, contact information, text messages, etc. all of which may be subject to privacy expectations. The algorithm applied to the data did not discriminate in what it was computing. If people have a reasonable expectation of privacy regarding the data on a cell phone, does that expectation extend to the hash values of the data?

The pending decision by the Supreme Court regarding the constitutionality of cell phone searches incident to an arrest (p. 64) will also have implications beyond just the specific cases being examined by the Court. While the Court's ruling will specifically address cell phone searches after someone is legally arrested, the decision will affect what law enforcement are allowed to do regarding all searches of cell phones and portable electronic

devices without a warrant.

A ruling that warrantless cell phone searches at the time of arrest are permitted will expand the authority of law enforcement officials. This could allow the officers in the above scenario to search the phone the moment it is found in Dana's car. The scope of the search would need to be determined. Full forensic analysis would reveal the most to law enforcement but could invade privacy. A hashed-based search would give the analyst the ability to control what is being searched for with privacy in mind, but the use of specific techniques is usually not determined by the courts. Reasonable suspicion might be used to place some limitations on the actions of the officers by requiring some justification before officers can examine the device.

A ruling that cell phones cannot be searched without a warrant would result in the potential loss of evidence for law enforcement. At most, if law enforcement felt the destruction of evidence was a possibility, they would be allowed to seize the phone. After establishing probable cause, a warrant could be obtained and then the search conducted. In the scenario above this would likely render the sector hash image of the SD card inadmissible unless the government could prove that Dana's consent extended to include the forensic analysis done in the off-site lab.

Where does a sector-based hash search fall under current law?

The science behind sector-based hash searches is the same as that used in file-based hash searches, just at a finer level. The use of cryptographic hash algorithms as a tool in digital forensics is well established. While sector-based hash searches do have a higher false positive rate than file-based hash searches, the reason is understood, can be mitigated, and is offset by the advantaged gained in having the ability to find matches to parts of files that would have been missed before. The investigators would need to verify that the hash value matches are not from common data blocks. Sector-based hash searches would also need to meet the criteria set by the judiciary in *Daubert v Merrell Dow Pharmaceuticals* (p. 33) but that could be done with reference to existing published work.

In this scenario, the sector hash image of the cell phone SD card is brought to a lab for analysis. The search conducted by the analyst is designed to only find matches to the video of the murder obtained by police. *United States v Jacobsen* (p. 29) and *Illinois v*

Caballes (p. 37) are the basis upon which certain search techniques are authorized without a warrant. Field tests of substances to determine if they are drugs and the use of search dogs are all justified under this precedent. The key factor that allows their use without a warrant is that the result of the technique must reveal only whether the object of the test is illegal or not. The search for matches between the SD card hash values and the hash values of the murder video could meet this criteria. The search can be implemented to only alert the analyst to matches of hash values which are know to be from the video. If that is the case, neither a warrant nor an exception to the warrant rule would be required to conduct the search, as only matches for known illegal content would be found.

Judicial precedent on this matter is not clear. *United States v Crist* (p. 44) saw all the evidence of child pornography found on a computer suppressed because the court determined the scope of a private search was exceeded. In that case, a forensic analyst conducted a file-based hash search that alerted to matches of hash values from known child pornography. The court did not see a hash value search as a technique which only alerted to illegal content. Since the matches resulted in the police learning more than what had been previously learned from the private search, the court deemed that a warrant or an exception to the warrant requirement was needed. The same could apply to this scenario and a court could require a warrant for any hash-based search.

The evidence found

The examination in the lab produced 10 matches of hash values from the sectors on the SD card to hash values of blocks of data from the video. The matches found could be used as evidence to obtain a warrant or be used as part of testimony by an expert witness. *United States v Miknevich* (p. 49) set the precedent that file hash values can be used to establish probable cause for a warrant. Extending this to cover the sector hash matches to the video would be reasonable but the matches must be validated first. As was shown in research involving block hashes, some types of files are prone to having common blocks regardless of the information held in the file. The analyst would need to confirm that the matches are from distinct blocks. Ideally this would be accomplished by pruning the database of hashes of common data blocks from the video prior to the search but can also be checked for after the fact by confirming with the block hash database that no other files have a match for that block hash.

While these matches do seem to meet the requirements for probable cause, it is unlikely that the matches alone would be enough to justify a conviction. The standard for criminal conviction is that the case must be clear beyond a reasonable doubt; it might be possible for a defense to come up with several hypotheses as to why the data blocks were present that would establish doubt such as the matches were found using a new technology, a software or operating system bug, or they may have been intentionally placed there or been intentionally manufactured to match. The matches might prove to be circumstantial at best.

5.2 Border Crossing Search Leads to Discovery of Hard Drives

The second scenario explores a second use of block hashes. Whereas the previous case used block hashes to find trace evidence of a known file, in this scenario block hashes are combined with random sampling to enable rapid triage. This scenario also avoids the question of consent by relying on the broad authority that is granted to law enforcement officers executing searches at border crossings.

5.2.1 Scenario

Taylor is a United States citizen who lives near the U.S.–Mexico border in a suburb of San Diego, CA. Taylor usually makes a trip to Mexico about once a week. He spends a couple days there and returns to the United States. When he pulls up to the border crossing, a Customs and Border Protection agent notices a laptop bag in the back seat of his car. He is told to pull over in order to have a search of his vehicle conducted. The computer is inspected to see if it is stolen property and a search dog finds nothing suspicious, but a visual inspection of the trunk reveals several 4TB hard drives. Taylor is informed that his drives will be searched for illegal content.

Inspection of the drives reveals that they are formatted with the Windows NT file system (NTFS) and that each contains approximately 100 files with sizes ranging from 1GB to 100GB, each ending with the "`.tc`*" extension. These files appear to contain TrueCrypt [69] encrypted containers.*

Taylor claims to be unfamiliar with the contents of the drives and says that he

cannot provide a password to decrypt them. It is not feasible to attempt to crack the containers using a brute force attack. Instead, the agent conducting the search decides to employ random sampling on several of the drives. The agent will use a tool to randomly select sectors from the drive, hash those sectors, and compare the hash values to a database containing the hash values of other encrypted containers that have been previously found. The search produces several matches to files found on other hard drives associated with other cases involving child pornography and drug trafficking. Based on a match of the hash value of the encrypted sector, the border agent has Taylor held until he provides the decryption key or can explain how the encrypted files came into his possession. The agent also contacts the law enforcement agencies which provided the hashes of the encrypted data in the database to help determine his next action.

5.2.2 Analysis

Customs and Border Protection agents have been granted a significant amount of authority when it comes to border searches in the interest of national security [33]. Suspicion that something is wrong is not required for CBP agents to search a person or property. While the power authorized for doing this type of search is currently being challenged [70], as it is being viewed as an invasion of privacy, the current precedent allows federal agents broad latitude in searching for illegal materials and contraband.

The use of random sampling

Because CBP agents have finite resources, searches must be limited to a reasonable amount of time. Research has shown that random sampling can produce a greater than 99 percent chance of finding a match of 100 MB of target data during the course of a 10 minute search [18]. The alternative of making a copy of the hard drive to allow a search of the whole volume would take significantly more time even if the drives were not encrypted. Random sampling provides a real solution to the growing issues that come with large amounts of available storage. It provides a method for investigators to quickly classify digital media as relevant or not with a very low probability that data of substance would be missed.

Random sampling in combination with sector hashing has not been adopted by the forensic

community at large. It remains relatively untested and has not been published in peer-reviewed literature. This could lead to challenges based on *Daubert v Merrell Dow Pharmaceuticals* (p. 33). It is based, however, on two well-grounded concepts in computer science: cryptographic hash algorithms and pseudo-random number generation. When a match is found, it is as accurate as any other hash-based search. The risk is assumed by the investigators who take a chance that pertinent evidence of illegal activity may be missed in the interest of speed.

Forensic analysis at border crossings

Current precedent, as seen in *United States v Ickes* (p. 40) and *United States v Arnold* (p. 43), allows a search of electronic media a person has with them at a border crossing. Reasonable suspicion is required if any search technique will be particularly intrusive or destructive. The decision in *United States v Cotterman* (p. 50) equated intrusive search to forensic analysis. That is, the courts have decided that turning on a computer and browsing the files is not intrusive and does not require reasonable suspicion but that a complete forensic analysis is intrusive and does require reasonable suspicion.

To the author's knowledge, the courts have not stated specifically which techniques fall into which category. Any hash-based search is typically viewed as a forensic technique and is considered intrusive. In *United States v Cotterman* (p. 50) the analysis was considered intrusive because it involved the recovery of deleted data and the decryption of password-protected files. The court equated searching for deleted files not only to finding out what is in a suitcase, but being able to determine everything that had ever been in the suit case. A hash-based search using random sampling has the potential to find hash value matches of deleted data, so reasonable suspicion might be required.

At the same time, the hash values themselves reveal nothing about the data from a privacy standpoint and only alert the analyst because they match data that is evidence of illegal activity. As seen in *Illinois v Caballes* (p. 37), the use of narcotics sniffing dogs is a well established means of conducting a search for illegal activity with no requirement for reasonable suspicion or probable cause. A sector-based hash search with random sampling can be implemented to be a reasonably quick search at the border that will only alert to evidence of illegal activity; such a search could be argued to be similar to a search using drug sniffing dogs.

The evidence found

The forensic analysis produced matches meaning that copies of the same TrueCrypt container had been found previously. TrueCrypt operates in two modes: it can be employed on digital media were the whole volume can be encrypted including the boot sector; or it can create containers inside of a digital media. The containers are an encrypted file that can be mounted like digital media. As files are added to the container, TrueCrypt encrypts the data and stores it in the container. Assuming that a different key is being used, if a file is copied from a container and placed into another container, the encrypted data streams will be different. The whole container itself can be copied, though, as it looks like a file to the file system. If this is the case, finding sector matches would be possible, since the exact sectors would have been copied with the entire file.

The matching hash values may belong to containers that have been decrypted and may contain known illegal content. Alternatively, the sectors may belong to containers that have not been decrypted but are connected to other criminal cases. If the matches are from known illegal content, the matches could be used for reasonable suspicion or probable cause for further investigation. As was stated in the first scenario, hash value matches can be used to establish probable cause.

The use of encryption brings up several issues. Is the fact that Taylor is crossing the border with several encrypted hard drives suspicious in itself or just good information assurance? In certain instances, a person can be compelled to provide a blood sample for DNA testing or finger prints. The forced production of decryption keys is currently before several courts, but that question is beyond the scope of this thesis.

5.3 Warrant Search of Files Using Similarity Matching

The third scenario examines the use of a similarity function to find data that is similar to other evidence found on other computers during the course of an investigation. Law enforcement will execute a search warrant authorizing the search of the suspects digital media, but must insure their search methods do not exceed the scope of the search.

5.3.1 Scenario

John is a United States citizen and is suspected of being a member of a criminal organization that is engaging in credit card fraud. Law enforcement officials execute a search warrant on John's premises based off probable cause established with evidence that resulted from the arrest of an associate. The warrant specifically authorizes the search for electronic records and documents related to the illegal collection, replication, and distribution of credit card numbers, bank account numbers, and other personal information. John's personal computer and several terabyte hard drives are confiscated and sent to a forensics lab for analysis. The analyst is specifically asked to find any files, especially documents, that are consistent with those found on other computers that are part of the investigation. In line with recent legal opinions (and as a result of judicial intervention), the warrant specifically disallows a "general search" of every file on the drive.

As part of the investigation the analyst conducts a file-based hash search of the drives, comparing the file hash values of John's computer hard drive and the external hard drives to the hash values of files known to be associated with the operations of this criminal organization. The search for files by hash turns up nothing. Rather than give up, the analyst uses a new similarity matching tool. The similarity matching tool does a byte-level comparison of files from John's digital media to the electronic records and documents found on the computers of other suspected members of this organization.

The similarity matching tool produces several matches. The analyst performs a visual inspection of the matched files and determines that some are pertinent to the current case but that some of the matches contain evidence of criminal activity for an unrelated case.

5.3.2 Analysis

If a warrant search is authorized then law enforcement have met both probable cause and particularity requirements. Using similarity matching as part of the forensic analysis will find related data that doesn't have to be an exact match to the target data set while reducing analyst workload.

Using similarity matching as part of forensic analysis

Similarity matching algorithms seek to reduce the workload of the analyst. Applying automation to the process successfully would reduce the amount of sifting that the investigator must do when searching for related evidence that is not an exact match to previously found information. The human must still be in the loop and verify that when a matching algorithm finds two objects that the algorithm determines are similar, that the objects do have the correct context the analyst is looking for.

The verification would be important as it is unlikely that current similarity tools and methods, on their own, would stand up to the challenges outlined by the judiciary in *Daubert v Merrell Down Pharmaceuticals* (p. 33). While it has been published in peer-reviewed publications, similarity matching has not yet achieved general acceptance in its community and its error rate might cause concern as it is higher for hash-based searches. The higher error rate means that a similarity search would return many documents not associated with the case which would not be consistent with the warrant.

Documents may be determined by similarity algorithms to be "matches" because they contain the same or similar human-generated content, or they may be deemed similar because of apparently inconsequential matches—for example, the two documents use the same fonts. Yet even for apparently inconsequential matches, there may be underlying reasons that are of interest to law enforcement—the two documents may contain the same fonts because they came from the same template that was distributed within a specific organization. A concern for analysts, and what appears to be unanswered is: when similarity matching is used in real situations, how much material that an analyst would consider similar is being missed?

Fitting similarity matching into the current legal framework

Warrants do have several requirements that must be specified. In particular, the person or property to be searched and the items to be searched for must be specified. Warrants are not required to state how the search will be conducted. This holds true for searches involving computers, as well as can be seen in the decisions *United States v Brooks* (p. 38), *United States v Hill* (p. 41) and *United States v Giberson* (p. 45). Courts have recognized the difficulty of trying to find relevant information when that data is commingled on a device which could store data anywhere on the media. Thus, a search broadened by using

a similarity function might be acceptable.

At the same time, courts are concerned about people's expectation of privacy in regards to their data. The Ninth Circuit Court of Appeals stated, as part of the decision in *United States v Heckenkamp* (p. 42), that a person has an expectation of privacy regarding personal computers that is both legitimate and objectively reasonable. There is an increased risk with similarity matching that data which is not pertinent will be found similar and exposed to government review as part of the verification process. With hash-based searches, a match means that the data found is an exact match for something already know to exist with an extremely low chance that there could be a collision. Similarity matching does not enjoy that same certainty. By the very nature of the similarity algorithm, the search will find matches that are not exact, increasing the chance that unrelated information will be exposed.

The decision in *United States v Grimmett* (p. 40) specified that the search of digital media may be as reasonable as required in order to find items described in the warrant. The idea of similarity matching seems very reasonable, in fact it might be a way reduce the invasion of privacy that can occur during a forensic examination if it can produce matches of only similar content with a known error rate. On the other hand, context matters when it comes to finding similar information. It might be difficult for some to recognize a technique as reasonable when it attempts to translate streams of data into a probability they are similar when the actual meaning of the data is not examined.

The decision in *Kyllo v United States* (p. 37) could be used to challenge the use of similarity matching all together. While sector-based hash searches and random sampling are based on techniques already established in forensics, similarity matching is a new way to conduct a search. This would be a misapplication of the decision for two reasons. In that particular case, it was the surveillance without a warrant that was found to be in violation of the Fourth Amendment. With a warrant, the use of new technology is legal as long the evidence it produces is relevant to the investigation and meets the standards of *Daubert*. In addition, the analyst already has the ability to access all the data available on the media. While similarity searching is finding data in a new way, a way that is optimized so as to reduce workload and time requirements, this method is not exposing anything more than what the analyst already has access to.

Evidence of another crime

In this scenario, the analyst, while examining the matches produced by the similarity matching program, discovered evidence of another crime. Judicial precedent in this area is not very clear. The decision in *United States v Carey* (p. 35) saw evidence of another crime discovered during the warrant search of a computer suppressed. It was not the specific technique or program used that led to the decision; it was the actions of the investigator. Upon discovering evidence of different criminal activity, the officer gave up the search authorized in the warrant and sought all the evidence relevant to the newly discovered illegal activity. In *United States v Mann* (p. 48), the court of appeals decided that the use of a feature of a tool that would alert to evidence of another crime exceeded the scope of the warrant. In that same case, however, evidence of a different crime that was stumbled upon during the search for evidence authorized by the warrant was allowed under the plain view doctrine. With similarity searches, it is known that unrelated data will be exposed, it is just the nature of the algorithm. While the intent of the analyst is to only find evidence related to the case, that person knows that information not pertinent to the case will be found and viewed. In a manner similar to *Carey*, which saw the evidence suppressed because of the intent of the investigator, it would be possible for a court to suppress evidence of another crime when similarity matching is used because the analyst knows for certain that irrelevant data will be found regardless of intent. On the other hand, the decision in*Mann* may see the discovery allowed under the plain view doctrine.

At least one court is worried about how the plain view doctrine can reduce the effectiveness of warrants at protecting the privacy interests of the population regarding searches of digital media. A set of guidelines was established by the Ninth Circuit Court of Appeals in their decision in *United States v Comprehensive Drug Testing, Inc* (p. 47) regarding the search of data on digital media. Those guidelines recommended investigators should waive reliance on the plain view doctrine when examining data due to how extensive and commingled it is on digital media. In addition, the guidelines state that only information uncovered for which a warrant is authorized may be released to investigators. This precedent would see the evidence of another crime dismissed regardless of how it was found. The ruling in *CDT* was highly controversial, and the degree that its language will be adapted by other courts remains unclear.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6:
Conclusion

The judiciary has to balance the rights and privacy expectations of citizens with the general safety of the population. The decisions of the courts set precedence that establish limitations that law enforcement must comply with when conducting searches. In the course of investigations, digital forensics techniques are used as part of that search effort, so the people who use them must comply with the precedent set by the judiciary. The United States judiciary system only decides on the legality of a new methodology after it has been used in the context of an actual investigation and the evidence is brought to trial. Laws, statutes, and precedent therefore inherently lag behind the technology used in investigations. This thesis examined current federal law and applied it to the use of new digital forensics techniques that could be used during the course of a search for data. Using hypothetical scenarios, we examined the use of sector-based hash search, random sampling, and similarity matching in the context of current federal law in the course of a consent search, border search, and warrant search respectively.

6.1 Hash-based Searches

The role that hashing fulfills as part of an examination has yet to be fully exploited. While precedence has been established that allows hash value matches to help meet the standard of probable cause, there is a hesitation to fully implement it that comes from privacy concerns. Current rulings are inconsistent, but they trend to treat hash values like the original media wherein people have a reasonable expectation of privacy. This can be seen with both warrantless searches and warrant searches that find evidence of another crime. In both cases, it is very easy for the scope of a search to be exceeded. This limitation conflicts with other accepted methods, such as chemical field tests for drugs or narcotics sniffing dogs, that are allowed without a warrant because they can only reveal evidence of specific illegal activities.

Sector Hashing

It appears likely that sector hashing can be adopted by investigators relatively easily under the current framework. The technique works very similarly to and is based on the same

principals as file-based hash searches which are common in forensic examination today. The adoption of sector-based hash searches brings the possibility of finding more pertinent data. Data are currently being missed with file-based hash searches because blocks of the file may be missing or corrupted. Sector hashing will find matches in these cases as long as one distinctive data block of the original file is intact.

Random Sampling

As more and more data need to be examined, forensic analysts must find techniques that will reduce the amount of time necessary to search media. Investigators need a way to quickly triage digital media so that time can be focused on the specific devices of interest. Random sampling has the real potential to help law enforcement focus their searches by providing a way to triage digital media quickly and with high probability that the data being sought wont be missed. As it is based on sector-hashing, it can be easily incorporated into the existing forensic framework. While it won't replace an in-depth analysis, it will give examiners a way to determine which media may be worth spending more time and resources on when a large collection of data needs to be examined as part of an investigation.

6.2 Similarity Matching

Similarity matching has the potential to bring new capabilities to the forensic analyst. Being able to find data that are similar and not just an exact match to data of interest using automation will reduce the analyst workload while providing visibility on data that would have remained hidden using traditional hash-based techniques.

Acceptance by the forensic and law enforcement community depends on several factors. Similarity matching, because it is looking for data that are similar but not exact matches, seems to raise more cause for concern. Similarity matching appears to be more like a general search than exact matching and may find matches that are not restricted to a specific scope. Because criminals do not clearly label their illegal activity as such, several courts have held that forensic investigators are within their authority to examine every file on a subject's hard drive and only present information that they find that is relative to the case. If this holds, then similarity matching algorithms will be just another tool, able to be used or not as a case requires, but not subject to specific legal requirements.

6.3 Unanswered Questions and Future Work

This thesis highlighted many of the issues federal law enforcement and forensic analysts must address in the course of an investigation involving the search of digital media. While the techniques presented seek to make those searches faster while revealing relevant information that may have previously been missed, there are still many areas that require further research.

This thesis limited its scope to just federal law. An analysis similar to the one presented could be done that addresses international law, the Uniformed Code of Military Justice, state law, or local law. How do these forensic techniques fit into the framework of these laws? In addition, the analysis could be expanded to cover other types of searches and the impact all three forensic techniques can have on those searches.

Real world testing with the tools presented would provide tremendous insight into their effectiveness. A case could be investigated where two different examiners conduct an analysis independently. One would use traditional forensics techniques and the other would implement the new methods suggested in the research. Such a study could help to determine if these techniques are easily deployable and how well do they perform compared to their traditional counterparts.

Forensic analysis primarily consists of evidence collectors seizing digital media and then making copies of the media for analysis back in a lab. Being able to deploy the techniques employed in a lab via remote access is an area that could be explored further. In the third scenario presented in this thesis, law enforcement seized a computer and several external hard drives. If the subject of the warrant had been using cloud services to conduct his illegal activity, how would the investigators have proceeded? Can forensic analysis be done remotely over an Internet connection? Can the company offering the cloud services be forced to provide access or make copies of the users data? In a distributed environment, how does the judiciary ensure privacy interests are protected?

The judiciary could eventually decide that people do not have a reasonable expectation of privacy regarding the hashes of their data. Such a decision could lead to broadly expanded authority for law enforcement: for example, the systematic searching of all cloud-based data. What would be the extent to which they could search for evidence of illegal activ-

ities with or without a warrant? Would investigators be permitted to collect and store the hash values of anyone that government agents had access to regardless of any evidence or suspicion?

In summary, while these three hypothetical scenarios have been a useful tool for investigating the legal ramifications of sector hashing, random sampling, and similarity searches, they are no substitute for the actual experience that will come from the application of these technologies to real cases.

REFERENCES

[1] W. Stallings and L. Brown, *Computer Security: Principles and Practices*, 2nd ed. Prentice Hall, 2011.

[2] R. Rivest. (1992, April). The MD5 message-digest algoritm. [Online]. Available: http://tools.ietf.org/html/rfc1321.

[3] *Secure Hash Standard,* FIPS180-4, 2012.

[4] M. Bellare and P. Rogaway, "Random Oracles are Practical: A Paradigm for Designing Efficient Protocols," in *Proc. 1st ACM Conf. on Computers and Communications Security,* Fairfax, VA 1993, pp. 62–73.

[5] E. Thompson, "Md5 collisions and the impact on computer forensics," *Digital Investigation*, vol. 2, no. 1, pp. 36–40, 2005.

[6] X. Wang and H. Yu, "How to Break MD5 and Other Hash Functions," in *Advances in Cryptology–EUROCRYPT 2005,* Aarhus, Denmark 2005, pp. 19–35.

[7] B. Carrier. (2014, January). The sleuth kit 4.1.3. [Online]. Available: http://www.sleuthkit.org.

[8] AccessData. (2013, June). Forensics toolkit 5. [Online]. Available: http://www.accessdata.com/products/digital-forensics/ftk.

[9] Guidance Software. (2014, January). EnCase. [Online]. Available: https://www.guidancesoftware.com.

[10] S. Mead, "Unique file identification in the national software reference library," *Digital Investigation*, vol. 3, no. 3, pp. 138–150, 2006.

[11] Federal Bureau of Investigation. (2003, May). Privacy impact assessment Child Victim Identification Program innocent names national initiative. [Online]. Available: http://www.fbi.gov/foia/privacy-impact-assessments/cvip.

[12] J. Young, K. Foster, S. Garfinkel, and K. Fairbanks, "Distinct sector hashes for target file detection," *Computer*, vol. 45, no. 12, pp. 28–35, 2012.

[13] S. Garfinkel, P. Farrell, V. Roussev, and G. Dinolt, "Bringing science to digital forensics with standardized forensic corpora," *Digital Investigation*, vol. 6, pp. S2–S11, 2009.

[14] D. Quist. (2012, July). State of offensive computing. [Online]. Available: http://www.offensivecomputing.net.

[15] National Institute of Standards and Technology. (2014, March). National Software Reference Library Reference Data Set. [Online]. Available: http://www.nsrl.nist.gov.

[16] J. Devore, *Probability and Statistics for Engineering and the Sciences*, 5th ed. Pacific Grove, CA: Brooks Cole, 2009.

[17] S. Garfinkel, A. Nelson, D. White, and V. Roussev, "Using purpose-built functions and block hashes to enable small block and sub-file forensics," *Digital Investigation*, vol. 7, pp. S13–S23, 2010.

[18] J. Taguchi, "Optimal Sector Sampling for Drive Triage," M.S. thesis, Dept. Comp. Sci., NPS, Monterey, CA, 2013.

[19] V. Roussev, "An evaluation of forensic similarity hashes," *Digital Investigation*, vol. 8, pp. S34–S41, 2011.

[20] J. Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation*, vol. 3, pp. 91–97, 2006.

[21] J. Kornblum. (2013, July). SSDeep version 2.10. [Online]. Available: http://ssdeep.sourceforge.net.

[22] V. Roussev. (2013, October). SDHash version 3.4. [Online]. Available: http://roussev.net/sdhash/sdhash.html.

[23] C. Shields. (2013, December). Welcome to similarity digest text (SDText) tool. [Online]. Available: http://www.cs.georgetown.edu/ clay/research/sdtext.html.

[24] V. Roussev, "Data Fingerprinting with Similarity Digests," in *Advances in Digital Forensics VI.* Hong Kong, China 2010, pp. 207–226.

[25] F. Breitinger and H. Baier, "Performance Issues about Context Triggered Piecewise Hashing," in *Digital Forensics and Cyber Crime,* Dublin, Ireland 2012, pp. 141–155.

[26] F. Breitinger et al., "Security and Implementation Analysis of the Similarity Digest SDHash," in *First International Baltic Conf. Network Security and Forensics,* Tartu, Estonia 2012, pp. 25–40.

[27] Computer Crime and Intellectual Property Section, Criminal Division. (2009, July). Searching and seizing computers and obtaining electronic evidence in criminal investigations. [Online]. Available: http://www.justice.gov/criminal/cybercrime/docs/ssmanual2009.pdf.

[28] *Almeida-Sanchez v United States,* 413 U.S. 266 (1973).

[29] Constitution of the United States of America Amendment IV. [Online]. Available: http://www.archives.gov/exhibits/charters/bill_of_rights_transcript.html.

[30] O. Kerr, "Searches and seizures in a digital world," *Harvard Law Review*, vol. 119, no. 2, pp. 531–585, 2005.

[31] Federal Rules of Criminal Procedure 41.

[32] Federal Rules of Evidence 101 – 1103.

[33] 19 U.S.C. §482.

[34] 18 U.S.C. §2510-2522.

[35] 18 U.S.C. §3121-3127.

[36] 18 U.S.C. §2701-2712.

[37] *Katz v United States,* 389 U.S. 347 (1967).

[38] *United States v Heldt,* 668 F.2d 1238 (DC Circuit 1981).

[39] *United States v Tamura,* 694 F.2d 591 (9th Circuit 1982).

[40] *United States v Jacobsen,* 466 U.S. 109 (1984).

[41] *Arizona v Hicks,* 480 U.S. 321 (1987).

[42] *California v Greenwood,* 486 U.S. 35 (1988).

[43] *Florida v Jimeno,* 500 U.S. 248 (1991).

[44] *Daubert v Merrell Down Pharmaceuticals,* 509 U.S. 579 (1993).

[45] *Frye v United States,* 293 F. 1013 (DC Circuit 1923).

[46] *United States v Carey,* 172 F.3d 1268 (10th Circuit 1999).

[47] *United States v Upham,* 168 F.3d 532 (1st Circuit 1999).

[48] *Kyllo V United States,* 533 U.S. 27 (2001).

[49] *Illinois v Caballes,* 543 U.S. 405 (2005).

[50] *United States v Brooks,* 427 F.3d 1246 (10th Circuit 2005).

[51] *United States v Ickes,* 393 F.3d 501 (4th Circuit 2005).

[52] *United States v Grimmett,* 439 F.3d 1263 (10th Circuit 2006).

[53] *United States v Hill,* 459 F.3d 966 (9th Circuit 2006).

[54] *United States v Heckenkamp,* 482 F.3d 1142 (9th Circuit 2007).

[55] *United States v Arnold,* 533 F.3d 1003 (9th Circuit 2008).

[56] *United States v Cartier,* 543 F.3d 442 (8th Circuit 2008).

[57] *United States v Crist,* 627 F.Supp.2d 575 (DC Middle PA 2008).

[58] *United States v Giberson,* 527 F.3d 882 (9th Circuit 2008).

[59] *United States v Comprehensive Drug Testing, Inc,* 579 F.3d 989 (9th Circuit 2009).

[60] *United States v Mann,* 592 F.3d 779 (7th Circuit 2010).

[61] *United States v Miknevich,* 638 F.3d 178 (3rd Circuit 2011).

[62] *United States v Cotterman,* 709 F.3d 952 (9th Circuit 2013).

[63] R. Salgado, "Fourth amendment search and the power of the hash," *Harvard Law Review*, vol. 119, no. 38, pp. 38–46, 2005.

[64] R. Burrows, "Judicial confusion and the digital drug dog sniff: Pragmatic solutions permitting warrantless hashing of known illegal files," *George Mason Law Review*, vol. 19, no. 1, pp. 255–290, 2011.

[65] J. Goldfoot, "The physical computer and the fourth amendment," *Berkeley Journal of Criminal Law*, vol. 16, no. 1, pp. 112–167, 2011.

[66] Brief for Petitioner, Riley v California, No. 13-132 (U.S. Mar. 3, 2014).

[67] Brief for the United States, U.S. v Wurie, No. 13-212 (U.S. Mar. 3, 2014).

[68] Brief of Center For Democracy and Technology and Electronic Frontier Foundation as Amici Curiae in Support of Petitioner in No. 13-132 and Respondant in No. 13-212, Riley v California, No. 13-132, U.S. v Wurie, No. 13-212 (U.S. Mar. 3, 2014).

[69] TrueCrypt Foundation. (2014, February). TrueCrypt 7.1. [Online]. Available: http://www.truecrypt.org.

[70] American Civil Liberties Union. (2013, December). Court rules no suspicion needed for laptop searches at border. [Online]. Available: https://www.aclu.org/national-security-technology-and-liberty/court-rules-no-suspicion-needed-laptop-searches-border.

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California